THIS WEEK

EDITORIALS

MISCONDUCT Japan sets out its strategy to tackle research fraud **p.8**

WORLD VIEW The moral and scientific case for a ban on seal products **p.9**

SOUTH SUDAN Conflict triggers food crisis and fears of famine **p.12**

No magic fix for carbon

Carbon capture and storage projects promise to make a dent in global emissions — but only as part of a broader programme of technology deployment and economic incentives.

The international pantomime that is climate-change politics is filled with heroes and villains, who jump onto and off the stage and trade places as time passes and the focus of attention changes. But one character endures: the fairy godmother, a single brilliant idea or advanced technology who with a single wave of her wand can introduce some magic to save the planet. It is a seductive and appealing plot twist, partly because it guarantees a happy ending, and partly because that happy ending comes about without any serious sacrifice by the dramatis personae. This *deus ex machina* principle of screen-writing — plot the hero or the world into a seemingly impossible corner and have the solution appear from nowhere in a puff of inspired smoke — infuriates science-fiction fans everywhere.

Over the past decade or so, carbon capture and storage (CCS) has been the fairy godmother of climate change, or at least of the politicians who have pledged in ever more ambitious terms to tackle the problem. Dig into most political promises to slash greenhouse-gas emissions by headline amounts — 80% by 2050, that kind of thing — and there she is. A significant proportion of the promised cuts are the result not of declines in carbon dioxide production, but of attempts to trap damaging emissions at source and divert them under the ground rather than into the atmosphere. Clean coal, CCS technology, capture-ready: the idea has spawned its own subplots and terminology. Regulations on carbon pollution permitted from new fossil-fuel-fired power plants are also being drawn up, on the assumption that CCS is feasible, and that it can be implemented on a massive scale.

Some of this political ambition has been backed with public investment. According to the International Energy Agency (IEA), from 2007 to 2012 more than US$12 billion of public funds around the world were made available to projects to demonstrate that the concept could work. Impressive perhaps, but hardly sufficient. The IEA has also said that to make the promised contributions to emissions targets, by the middle of this century CO_2 storage would have to be a well-developed industry in its own right — bigger than last year's global oil and gas industry, with all of the associated infrastructure. About 25 million tonnes of carbon dioxide are already piped under the ground each year for a variety of reasons. The IEA says that must rise to 7 billion tonnes by 2050.

As we report on page 20, two coal-fired power plants in North America are preparing to nudge up the modest annual amount of CO_2 sequestered. The Boundary Dam Power Station in Saskatchewan, Canada, will probably be first. It is scheduled to switch on later this year, and if it does so it will win a global race. For the first time, a commercial-scale plant that supplies electricity to the grid will capture and store most of its emissions, about 1 million tonnes of CO_2 a year. (Whether this is a good thing for the environment depends on your point of view: the gas will be sold to an oil company and squeezed underground to help to flush out the stubborn reserves of an oilfield.)

Following close behind is a more modern coal plant in Kemper County, Mississippi, designed to capture 3.5 million tonnes of CO_2 a year — about two-thirds of its total emissions. This captured gas will also go towards enhanced oil recovery when the plant starts to operate towards the end of this year.

The concept works. The question is, at what cost? As Howard Herzog, a CCS researcher at the Massachusetts Institute of Technology in Cambridge, says in the News story: "The technology is ready to go. The problem is that policies aren't in place to make projects economic." Well, quite.

> ***"Carbon capture and storage has been the fairy godmother of climate change, or at least of politicians."***

The commercial market for CO_2 is small and unlikely to expand any time soon. Schemes to make companies pay for their emissions were intended to penalize polluters and level the playing field for clean but pricey alternatives, but they are struggling. However cheap CCS technology might get, a coal or gas plant that scrubs its exhaust gases to capture the carbon will always be more expensive to run than one that does not — making it the first to be turned off when demand for electricity falls outside peak times.

Many questions remain about the long-term viability of a serious and sustained CCS contribution to the global effort to reduce greenhouse-gas emissions, not least how to guarantee that stored carbon stays stored. But by this time next year, the coal plants in Saskatchewan and Mississippi could give politicians around the world sufficient proof that the concept can be deployed — not as a fairy godmother to spirit away their problems, but as part of a broader suite of technologies. Then they just have to decide what to wish for. ■

False positives

A correlation between error rate and success undermines promise of stem-cell trials.

When it comes to stem-cell therapies, the stakes are high — but not as high as the hopes of people who are severely ill. Over the past few years, dozens of small, early-phase clinical trials have tested the value of adult stem cells in treating debilitating or life-threatening heart disease. Results have been mixed, but most peer-reviewed academic reports have hinted that patients may be helped. This has, understandably, encouraged clinicians to move potential therapies into large and expensive phase III trials to establish whether the treatments can fulfil their promise.

Now comes a shocking reality check, revealed this week in the *British Medical Journal* (*BMJ*). As we report on page 15, a London-based team has scrutinized reports of all the randomized trials of bone-marrow stem-cell treatments for heart disease they could find.

The authors searched for discrepancies that might undermine the results and found plenty — errors such as numbers not adding up, or individual patients reported variously as male and female, dead and alive. In fact, the researchers found a linear relationship between the number of discrepancies and the claimed effect size. The small number of trials that they identified as unflawed showed an effect size of zero. In other words, the scientists declare this stem-cell emperor to have no clothes.

The multitude of discrepancies may not necessarily invalidate the conclusions of an individual trial — the authors point out that all too often the clinical data are not available, leaving them unable to check whether the discrepancies are real errors or just the result of sloppy reporting.

But, at the very least, the *BMJ* report should raise the question of whether the data are really strong enough to support the big step of moving to a phase III trial, particularly given that in the case of adult stem cells the results of animal studies have been ambiguous. Initially, researchers suggested that these cells became specialized to the target organ and replaced damaged tissue, but this idea has since been rejected. Many clinicians now think that the cells instead act to heal the surrounding tissue, releasing molecules that cause inflammation and the growth of oxygen-bearing small blood vessels, processes important to repair.

The findings of the *BMJ* study raise another worrying question: why did the clinical journals concerned fail to notice the discrepancies, given that many of the errors seem, in hindsight at least, to be startlingly visible? If a table claims to describe n clinical events, for example, but in its columns refers to $n+2$ events, is that really so hard to catch?

This, in turn, raises more queries about the process. Who should take responsibility for fact-checking a paper for internal consistency? Is it the notoriously busy clinical experts who act as referees? Or the editors, many of whom also have a full schedule of clinical duties? Few of the journals that published the papers scrutinized in this case have professional editors or significant numbers of in-house editing staff. Pressure to review and publish quickly is high. The two sides of the equation don't balance, and the problems identified in the study suggest something of a crisis.

"The small number of trials identified as unflawed showed an effect size of zero."

To address this, the publishers of clinical journals must do more to ensure that someone takes responsibility for the fact-checking. That could involve asking authors to guarantee that they have checked figures, tables, text and abstracts for internal consistency. Publishers could require authors to make available suitably anonymized data on each patient as metadata to the study, so that readers can trace the source of any discrepancy that might creep through. Or the publishers could reach into their pockets and provide more in-house resources to perform the necessary checking. What is not acceptable is for the situation to continue as it is, with responsibilities undefined and inexact publishing distorting clinical messages.

The problem seems to run deeper than the heart and stem-cell studies checked in this case. For years, analyses have highlighted a bias towards publishing clinical trials that show a positive outcome. (A similar trend has also been found with scientific results.)

Translational medicine is one of the buzz-phrases of the twenty-first century. In a way, it should be a surprise that it has taken so long for the idea to catch on. What use is medicine that is stuck in the scientific laboratory? But as the curious case of adult stem cells demonstrates, the right checks and balances are not brakes on progress, but an essential foundation for that progress. Fools rush in. So do those who have not done their homework. ■

Agency for change

Japan's proposed reforms to science monitoring are welcome but long overdue.

Scientific misconduct is a universal problem. Policies to investigate and prevent it, however, are patchy. Japan is now taking welcome steps to address the issue.

Japan has certainly produced some of the more bizarre cases of scientific fraud identified in recent years. In 2000, an amateur archaeologist was caught on film burying stone tools that he later unearthed as evidence of human civilization — his 'discoveries' over two decades falsely pushed back Japanese history by 650,000 years and corrupted a generation of history textbooks (see *Nature* **408,** 280; 2000).

In 2009, a University of Tokyo professor, Serkan Anilir, was found to have lied about several of his career achievements, including his claim to be the first Turk in a NASA programme: an image of him wearing a spacesuit was uncovered as a fake. And in 2012, the 20-year career of an anaesthesiologist came under question amid the record retraction of more than 100 of his papers (see *Nature* **489,** 346–347; 2012).

There is more to these cases than embarrassing tales of individuals gone off the rails. They indicate a lack of oversight in research and the common cultural reluctance of colleagues to act on suspicions for fear of challenging their peers. They highlight how misconduct is not reported enough in Japan, partly because the country has lacked a high-level agency to deal with it.

Japan is now preparing to clean up its scientific act. At a 14 April meeting of the Council for Science and Technology Policy (CSTP), the nation's highest science-policy organization, an eight-person subcommittee called for the cultivation of research integrity in individual researchers, and for the setting up of fraud prevention and response measures at the institutional level to restore public faith in science.

The council's chair, Japanese Prime Minister Shinzo Abe, expressed concern that "the recent rash of cases involving scientific misconduct threatens to erode the foundation of our research". He noted that an approach to misconduct based purely on the experience of individual cases is inadequate; instead, he has asked the CSTP to develop measures "from a broad perspective".

In its call for action, the CSTP cited the ongoing case of Haruko Obokata of the RIKEN Center for Developmental Biology in Kobe. In January, she published research in this journal that suggested adult cells can be reprogrammed into stem cells through stress. Within weeks, allegations emerged that the work contained errors. On 1 April, RIKEN charged Obokata with misconduct. She is appealing the decision.

It is unclear how Japan will act on the CSTP call for action, but the country should take this opportunity to create an agency, akin to the US Office for Research Integrity, that can handle allegations of fraud and misconduct in a systematic way and encourage whistleblowers to come forward. The need for such an agency has been noted often, including in these pages (see *Nature* **437,** 595–596; 2005).

Researchers now deal with more data than ever before, and the evaluation of misconduct allegations often comes down to distinguishing sloppiness from deception in the presentation of data. For this reason, Japanese institutions should be given funding to educate their researchers in the responsibilities of data management. Whatever the outcome of the CSTP's proposals, the high level of attention given to the issue is long overdue. ■

➲ NATURE.COM
To comment online, click on Editorials at: go.nature.com/xhunqv


EVA BUTTERWORTH

The moral problem with commercial seal hunting

The Canadian seal hunt leads to animal suffering, and a European Union ban on the import of its products should stand, says **Andy Butterworth**.

As I write this piece in mid-April, small boats are weaving their way through the shifting Arctic sea ice off the east coast of Canada to hunt seals. In previous years, more than 1,000 boats have taken part. By the end of the 2014 hunt, a quota of some 400,000 young seals could have been taken, leaving the ice stained red with blood. The pace is rapid. Most of the annual take occurs in the first five days.

The annual Canadian commercial seal hunt is the world's largest hunt of marine mammals. A few weeks old, the seal pups are prized primarily for their skins and also for the omega-3-rich oil used in food supplements — products that are shipped around the world.

This month, the World Trade Organization (WTO) in Geneva, Switzerland, is expected to announce whether products from commercial seal hunting can be marketed in Europe. At present, they cannot. Such products have been banned by the European Union (EU) since 2009 to protect 'public morals'. Canada and Norway have asked the WTO to overturn the ban — the first of its kind — and the trade body will soon deliver its verdict.

My research has been at the centre of the case, and papers I published on welfare aspects of the seal hunt in the journal *Marine Policy* (see, for example, A. Butterworth and M. Richardson *Mar. Policy* **38,** 457–469; 2013) were cited more than 40 times in a WTO panel report last year that backed the European ban. I have been part of the EU delegation to the WTO, and had a ringside seat to view an international tussle between the promotion of trade, animal welfare and public morals. The WTO should reject the appeal, and the ban on seal products must stay in place. Here is why.

IT IS **IMPOSSIBLE** TO MONITOR ALL THE **HUNTING** ACTIVITY.

As an official observer, I have seen the hunt at close quarters, from the ice and from helicopters. The details are grisly. That is why the WTO originally agreed that the EU could act to limit trade on the grounds of public morals — the first time that such a restriction had been put in place.

When they are born, seal pups have white fur. They are suckled, weaned and then abandoned by their mothers at about 12 days of age. Stranded on the unstable ice, they remain alone and unfed for up to six weeks, and during this time their fur changes from white to grey — and the hunters arrive.

The pups are either shot from boats, or clubbed with a wooden bat or an iron-tipped pole called a hakapik. Some shot and injured seals slide into the water and are lost. Many shot and injured animals could potentially suffer for several minutes while the hunters manoeuvre their boats close enough to club them unconscious. If the ice is too unstable for the hunters to cross, shot and injured but conscious and reactive seals can be dragged into the boats with long hooked gaffs before being clubbed.

➲ **NATURE.COM**
Discuss this article online at:
go.nature.com/n9cfqi

My publications on the hunt reflect my scientific and veterinary perspective. Animal welfare can be scrutinized and measured. I brought both my compassionate (veterinary) and dispassionate (scientific) perceptions to bear. It is impossible to monitor all the hunting activity, which takes place across a huge shifting frozen expanse the size of France. But what I have seen and measured on the ice challenged me, and provoked concern within me. The Canadian Marine Mammal regulations allow hunters to retrieve injured animals from the ice using the hooked gaffs before they have been checked for unconsciousness. But people do not allow land animals to be treated in this way. If what I have witnessed being done to a young seal was done to a horse or a dog, there is little doubt that it would be labelled as cruel.

The EU ban was introduced on moral grounds, and science and scientific evidence can inform judgements on moral questions. When it comes to the seal hunt, the science indicated that some shot seals took a considerable period of time to die, and some injured animals were 'unchecked' for periods of several minutes before being finally killed by clubbing. The post mortems that we carried out on the ice indicated that some seals had multiple shooting, clubbing and hooking injuries — and that some had swallowed their own fresh blood — probably indicating that they were alive for a period following the first contact with the hunter. The assessments also described the distressed behaviour of conscious injured animals in response to being recovered from the ice with the gaffs.

Interestingly, the appeal from Canada and Norway does not challenge the "poor welfare outcomes" of the seals, which the WTO last year judged sufficient to justify the European ban. Instead, the appeal concentrates on trade issues and claimed unfair restrictions.

In reaching its decision, which will be final and binding, the WTO must reconcile contrasting statements from international agreements that are almost 70 years old. One forbids "arbitrary or unjustifiable discrimination" between countries. Another says that nations can act in a way that is "necessary to protect public morals".

As a human and as a veterinary scientist I consider the hunt to present real and significant welfare concerns. The available scientific evidence supports that opinion. But science, of course, is only one of the factors at play. Perhaps the final word should go to a statement attributed to Mahatma Gandhi: "The greatness of a nation and its moral progress can be judged by the way its animals are treated." ■

Andy Butterworth *is a senior lecturer and researcher in clinical veterinary science at the University of Bristol, UK.*
e-mail: andy.butterworth@bristol.ac.uk

RESEARCH HIGHLIGHTS

Selections from the scientific literature

CLIMATE CHANGE

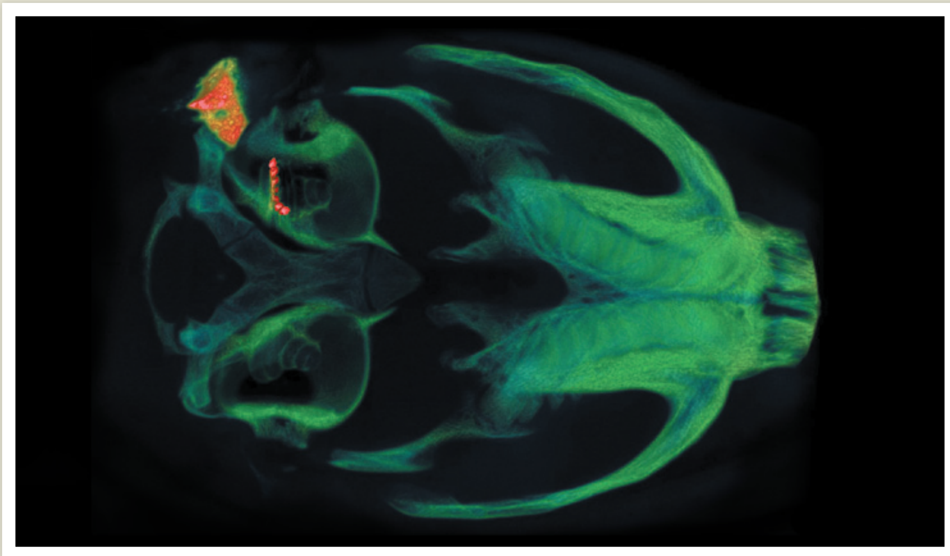
California drought linked to humans

Man-made global warming seems to have contributed to the extreme Californian drought in recent months.

Simon Wang of Utah State University in Logan and his colleagues re-analysed meteorological records and found that the ongoing dry spell is linked to changes in large-scale atmospheric pressure and circulation patterns. These typically occur over North America in the years before El Niño events — the occasional warming of the tropical Pacific Ocean.

Simulations show that atmospheric conditions that tend to cause intense drought in California have become more frequent since 1970 as a result of increased greenhouse-gas concentrations. Projections suggest that future droughts in California will be even more severe, the authors warn.

Geophys. Res. Lett. **http://doi.org/sfs (2014)**



T.-T. HUNG AND A. KWEK/UNSW-BRIL & ANIF; J. PINYON & G. HOUSLEY/UNSW-TNF

GENE THERAPY

Ear implant helps to repair nerves

Researchers have repaired an auditory nerve by introducing a gene into ear cells — after zapping them with electricity from a hearing device.

Electrical impulses from cochlear implants not only stimulate the auditory nerve in people with deafness, but also make cells permeable to DNA.

Gary Housley and his colleagues at the University of New South Wales in Sydney, Australia, used cochlear implants in deaf guinea pigs (pictured; implant in red) to deliver a gene encoding neurotrophin, a protein that stimulates nerve growth, to specific inner-ear cells. They found that the auditory nerve began to regenerate, extended towards the cochlea and showed greater sensitivity than in untreated animals.

A similar method using electrodes implanted in the brain could repair brain cells as a way to treat neurological disorders, the authors say.

Science Transl. Med. **6, 233ra54 (2014)**

For a longer story on this research, see go.nature.com/y3bxzr

PLANT BIOLOGY

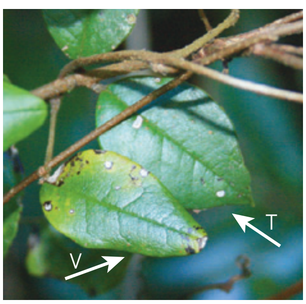
Leafy master of disguise

A climbing vine can mimic the leaves of any of a dozen host trees, possibly helping it to resist predation by herbivores.

Ernesto Gianoli and Fernando Carrasco-Urra of the University of Concepción in Chile looked at 45 samples of the vine *Boquila trifoliolata* (**pictured**, left), which lives in the temperate South American rainforest. The duo found that vines adapt 9 leaf features, including size, shape and colour, to mimic the leaves of any of 12 different host species (pictured, *Rhaphithamnus spinosus*; right).

Other plants are capable of mimicry, but no plant has been described before that can change its appearance to match that of so many different hosts.

Curr. Biol. **http://doi.org/sjg (2014)**



OSCAR GODOY

PALAEONTOLOGY

Flying reptiles were land lovers

Researchers have discovered the oldest fossil of a pterodactyloid, the group that includes the largest-known flying animals. The finding suggests that the ancient reptiles that gave rise to these creatures originated on land, rather than in marine environments.

The pterodactyloids evolved from smaller pterosaurs, the remains of which have been found predominantly in ancient seas. A team led by Brian Andres, at the University of South Florida in Tampa, analysed a roughly 163-million-year-old fossil from northwest China. They conclude that the animal, which had a 1.4-metre wingspan, is the earliest known pterodactyloid, on the basis of features such as an elongated metacarpus wing bone.

By comparing the new species — which the team has provisionally named *Kryptodrakon progenitor* — with its relatives and with modern flying vertebrates,

the authors suggest that the creature's ancestors evolved primarily on land.
Curr. Biol. **http://doi.org/sjf (2014)**

ECOLOGY

Chernobyl birds adapt to radiation

Researchers studying birds near the Chernobyl nuclear-disaster site in Ukraine have found the first evidence of wild animals adapting to ionizing radiation.

Ismael Galván, now at the Doñana Biological Station in Seville, Spain, and his colleagues examined feather and blood samples from 16 bird species, including barn swallows (*Hirundo rustica*) and wood warblers (*Phylloscopus sibilatrix*), in and around the Chernobyl exclusion zone.

They found that birds from more-contaminated areas had higher levels of antioxidants — which mop up damaging free radicals produced by radiation exposure — than birds from areas that were less contaminated. Animals from high-radiation regions also had better body condition and showed decreased DNA damage and oxidative stress.
Funct. Ecol. **http://doi.org/sh7 (2014)**

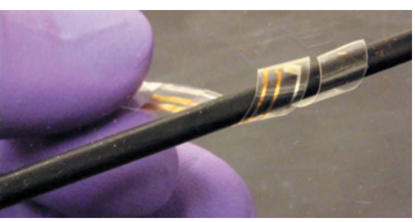
MATERIALS

Electronics mould to body tissue

Shape-shifting polymers could one day be used in implantable electronic sensors that conform to tissue inside the body.

A team led by Takao Someya at the University of Tokyo attached organic thin-film transistors to polymers that deform when warmed. They first fabricated the device in the shape of a helix and then flattened it. When they applied heat, the device wrapped itself around a thin cylinder (**pictured**). Under the skin of a rat, the film softened to follow the contours of the tissue 24 hours after implantation, while still maintaining its electronic properties.

The authors say that their system offers advantages over other materials: stiff substances cannot move with the body and elastic materials are difficult to insert *in vivo* and hard to produce in bulk.
Adv. Mater. **http://doi.org/sg7 (2014)**



TAKAO SOMEYA AND WALTER VOIT

MICROBIOLOGY

Immune system boosts microbe

A protein secreted by human immune cells prompts an infectious bacterium to mutate, protecting it from drugs and immune attack.

Pseudomonas aeruginosa is a leading cause of fatal lung infections in people with cystic fibrosis (CF). A team led by Daniel Wozniak of Ohio State University in Columbus found that LL-37, a human protein produced by immune cells called neutrophils, triggered the bacteria to make a sticky, sugary coating that is often seen in chronic infections of *P. aeruginosa*.

At low levels, LL-37 entered bacterial cells and bound to their DNA. By disrupting DNA replication, the protein induced mutations that boosted production of the sugar coating and resistance to the antibiotic rifampin. The mutations were similar to those seen in *P. aeruginosa* isolated from people with CF.
PLoS Pathog. 10, **e1004083 (2014)**

GENOMICS

Genome of the tsetse fly decoded

The genome sequence of the disease-carrying tsetse fly reveals potential targets for fly-control measures to stop the spread of disease.

Tsetse flies are the sole carriers of a protozoan parasite that causes trypanosomiasis in humans (known as sleeping sickness) and in livestock throughout sub-Saharan Africa. Members of the International *Glossina* Genome Initiative sequenced the DNA of the tsetse fly *Glossina morsitans morsitans*, which transmits the parasite to livestock. They found genes that are responsible for some of the fly's unique features, such as its ability to produce milk for its offspring and its attraction to the colour blue.

The genomic data could enable researchers to develop methods for manipulating the insect's reproduction, for example, thus helping to control tsetse-fly populations.
Science 344, **380–386 (2014)**
For a longer story on this research, see go.nature.com/yrum7w

SOCIAL SELECTION

US research censure stirs up Twitter

The US biomedical-research system is overcrowded, underfunded and headed for decline, according to a much-discussed report by four top-tier researchers, including Nobel prizewinner Harold Varmus, director of the US National Cancer Institute. The authors argue that the imbalance between the surge of talented scientists and the steady dwindling of funds for research has created a "hypercompetitive" atmosphere that stifles careers and hinders scientific progress. The paper resonated with young researchers, guaranteeing significant coverage on Twitter. John Bachman, a graduate student in a systems-biology lab at Harvard University, tweeted that the paper was a "must-read".
Proc. Natl Acad. Sci. USA 111, **5773–5777 (2014)**

Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

NATURE.COM
For more on popular papers:
go.nature.com/nd3ki1

CLIMATE CHANGE

Monsoon wet spells get wetter

The Indian subcontinent has seen greater extremes in rainfall during the South Asian summer monsoon season over the past 30 years than before 1980.

Nearly 20% of the world's population is affected by this monsoon, which brings most of India's annual rainfall. Deepti Singh and her colleagues at Stanford University in California compared precipitation data during wet and dry spells in two time periods, 1951–80 and 1981–2011.

Statistical analyses revealed that wet spells have grown more intense during the recent period, and that dry spells have become more frequent. The authors also found increased variability in daily rainfall over the past half-century, and an overall drop in peak-season precipitation.

These changes were associated with shifts in atmospheric circulation patterns and increased moisture availability. The more intense wet spells increase the risk of floods and the more frequent dry spells could negatively affect agriculture, the authors say.
Nature Clim. Change **http://dx.doi.org/10.1038/nclimate2208 (2014)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

ANDREEA CAMPEANU/REUTERS/CORBIS

Famine looms in South Sudan

Resumed fighting in South Sudan has made it impossible for many farmers to get hold of seed and tools, disrupting the main planting season and further exacerbating a food crisis. It also means that suitable land is scarce. Aid agencies said in a report published on 22 April that the food crisis is already causing worrying levels of malnutrition. United Nations secretary-general Ban Ki-moon has called for "immediate action" to avert a famine that could affect up to a million people. The last famine in Africa hit Somalia in 2011 and killed around 260,000 people.

RESEARCH

Probe head resigns

The head of a Japanese committee investigating misconduct in two stem-cell research papers resigned from the panel on 25 April. Shunsuke Ishii, a molecular biologist at the RIKEN Molecular Genetics Laboratory in Saitama, Japan, stepped down over anonymous allegations that at least one of his own papers contains problematic data. He says that he has acted out of concern that the incident could complicate the investigation. Ishii was leading a committee investigating research led by Haruko Obokata of the RIKEN Center for Developmental Biology in Kobe, which was published in *Nature*. See go.nature.com/t3i8md for more.

Middle East virus

As of 23 April, 345 people are reported to be infected with Middle East respiratory syndrome coronavirus, of whom 107 have died, according to the European Centre for Disease Prevention and Control in Stockholm. Most cases have occurred in Saudi Arabia, leading to the sacking last week of the country's health minister Abdullah al-Rabeeah. The virus, which was first identified in 2012, predominantly infects animals, but can spread to and between humans.

Stem-cell fraud

An Italian public prosecutor has criticized Davide Vannoni, president of the Stamina Foundation in Turin, Italy, for treating seriously ill patients with unproven stem-cell treatments. In a report published on 23 April, prosecutor Raffaele Guariniello concluded that Vannoni and 19 others, including physicians, civil servants and hospital directors, fraudulently obtained money — in some cases, tens of thousands of euros per patient — for the unproven treatments. A judge must now decide if the case should go to trial.

Dengue vaccine

An experimental vaccine gives children partial protection against dengue virus, drug manufacturer Sanofi Pasteur of Lyon, France, announced on 28 April. A large trial enrolled more than 10,000 children in southeast Asia; those who received the vaccine were 56% less likely to develop dengue than those given a placebo. More details, as well as results from a separate large-scale trial of the vaccine, are expected later this year. Dengue virus infects an estimated 100 million people worldwide each year, causing flu-like symptoms in most and a potentially deadly haemorrhagic fever in a small minority. No vaccine is approved to prevent infection.

FACILITIES

British ice-breaker

The United Kingdom will build a £200-million (US$336-million) polar ice-breaking ship by 2019, the government announced on 25 April. The British Antarctic Survey research vessel will be bigger and better-equipped than the nation's existing polar ships, which are scheduled to be decommissioned at the end of the decade. The government also launched a consultation into long-term UK plans for research-infrastructure spending; it intends to publish a spending road map by the end of the year. See go.nature.com/2ei5pl for more.

Waste-plant report

A nuclear-waste dump near Carlsbad, New Mexico, which suffered a radioactive leak in February, has been criticized for lax safety and dilapidated equipment. In a report released on 24 April, the US Department of Energy also said that the Waste Isolation Pilot Plant's response to the accident was slow and inadequate. The 665-metre deep plant is supposed to store casks of nuclear waste

for 10,000 years. But after just 15 years of operation, at least one cask ruptured and low levels of radioactivity leaked to the surface.

POLICY

E-cigarette rules

The US Food and Drug Administration (FDA) said on 24 April that it intends to expand its regulation of tobacco products to include cigars, hookah and pipe tobacco and electronic cigarettes. If the proposed rules go into force, these products would need to be reviewed by the FDA before being marketed, and they would not be allowed to be sold to people under the age of 18. The FDA will collect public comments on the proposal for 75 days, after which it may make a final ruling. See go.nature.com/lpwaw4 for more.

Starfish culled

Over the past two years, the Australian government has overseen a cull of some 250,000 crown-of-thorns starfish (*Acanthaster planci*; **pictured**), environment minister Greg Hunt announced on 22 April. The initiative used an efficient single-injection method to destroy the creatures, which pose one of the biggest threats to the Great Barrier Reef. The injection causes an allergic

GARY BELL/CORBIS

reaction that results in the starfish breaking apart and dying within 24 hours. It has produced a fourfold increase in deaths over previous methods, which used multiple injections. The cull is part of Australia's Reef 2050 Plan for the sustainable management of the reef between 2015 and 2050.

BUSINESS

Drug-sale record

Pharmaceutical company Gilead Sciences in Foster City, California, announced on 22 April record-breaking sales of its latest hepatitis C drug. Sovaldi (sofosbuvir) was launched last December and has since generated US$2.3 billion for the firm. The therapy has set a new record for drug sales; the previous record was held by another hepatitis C drug, developed by Vertex Pharmaceuticals of Boston, Massachusetts, that earned just over $1.5 billion in its first year on the market. Sovaldi is among the first treatments to fight the hepatitis C virus without requiring the use of interferon, a drug that causes unpleasant side effects.

Pharma swap

Two major pharmaceutical companies announced that they are overhauling their businesses on 22 April. GlaxoSmithKline, based in London, has sold its cancer-drug business to Swiss pharmaceutical giant Novartis, based in Basel, for up to US$16 billion, depending on the results of a clinical trial. Novartis, meanwhile, unloaded its vaccine programme, excluding flu vaccines, to GlaxoSmithKline for as much as $7 billion plus royalties. The two companies will combine their consumer-health programmes to create a new business.

TB drug trial

The first regimen designed to treat both drug-sensitive and resistant strains of tuberculosis will be tested in a phase III clinical trial. TB Alliance, a non-profit drug-development organization based in New York, said on 23 April that it hopes to launch the trial in four continents by the end of the year. The Shortening Treatments by Advancing Novel Drugs (STAND) trial will test a combination of three drugs called PaMZ (PA-824, moxifloxacin and pyrazinamide) that should shorten the time it takes to treat the disease from up to 24 months to just 6 months.

CRISPR claims

A biopharmaceutical company has raised US$25 million to develop a business focused on the acclaimed gene-editing technology CRISPR–Cas9. CRISPR Therapeutics of Basel, Switzerland, announced on 24 April that it had won funding from Versant Ventures, a venture-capital firm in Menlo Park, California. The development follows last week's announcement by the Broad Institute in Cambridge, Massachusetts, that it had patented an engineered CRISPR system (see *Nature* **508,** 436–437; 2014). CRISPR Therapeutics' claim on the technology's intellectual property could conflict with that of the Broad Institute.

COMING UP

3–7 MAY
The American Psychiatric Association meets in New York. Discussions will cover use of complementary and alternative medicine in psychiatry; treatment for military patients; and psychosomatic medicine.
go.nature.com/49cwbn

17–20 MAY
The American Society for Microbiology meets in Boston, Massachusetts. Topics to be discussed include microbial responses to climate change; microbes involved in metabolism; and the origins of zoonotic pathogens.
go.nature.com/2rpw4w

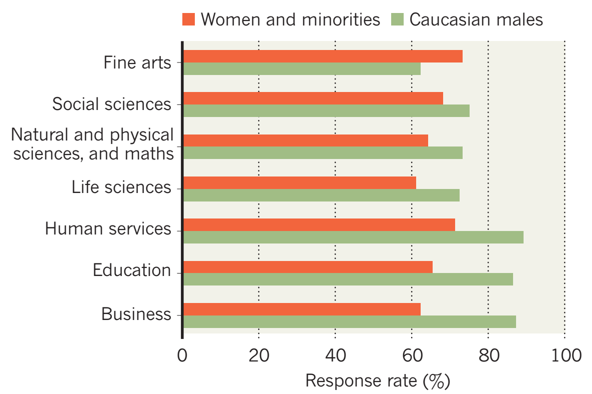
TREND WATCH

Women and people from ethnic minorities face well-documented hurdles in academia, including overt discrimination — and the bias extends to getting onto doctoral programmes. A 23 April study reports on sending e-mails to 6,548 professors at US institutions, in which (made-up) students requested a meeting to discuss research opportunities. It found that white men were more likely than women or ethnic minorities to receive a reply, in all but one discipline. See go.nature.com/1wjecs for more.

SOURCE: K. MILKMAN

BIASED TEACHINGS

Instructors at 259 US institutions were, on average, more likely to respond to fake e-mail requests for mentoring if the senders' names sounded white and male.



➲ NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

SECURITY Big money plans for 'vaccines on demand' split US researchers **p.16**

R&D Technology companies move to pick up slack in basic research funding **p.18**

ENERGY Canadian coal power plant takes carbon capture into the real world **p.20**

DISEASE The quest to isolate smallpox virus from ancient remains **p.22**

CHAIWAT SUBPRASOM/REUTERS/CORBIS

Many companies around the world offer stem-cell treatments to patients with heart disease.

THERAPEUTICS

Doubts over heart stem-cell therapy

Study queries early-phase trials of heart-disease treatment.

BY ALISON ABBOTT

An analysis of clinical studies that use adult stem cells to treat heart disease has raised questions about the value of a therapy that many consider inappropriately hyped.

Early-phase clinical trials have reported that adult stem cells are effective in treating heart attack and heart failure, and many companies are moving quickly to tap into this potentially lucrative market. But a comprehensive study that looked at discrepancies in trials investigating treatments that use patients' own stem cells, published this week in the journal *BMJ* (ref. 1), finds that only trials containing flaws, such as design or reporting errors, showed positive outcomes. Error-free trials showed no benefit at all.

The publication comes as two major clinical trials designed to conclusively test the treatment's efficacy are recruiting thousands of patients.

The *BMJ* paper "is concerning because the therapeutic approach is already being commercialized", argues stem-cell researcher Paolo Bianco at the Sapienza University of Rome. "Premature trials can create unrealistic hopes for patients, and divert resources from the necessary basic studies we need to design more appropriate treatments."

Therapies that use adult stem cells typically involve collecting mesenchymal stem cells from bone marrow taken from the patient's hip bone. The cells are then injected back into the patient, to help repair damaged tissue. Original claims that they differentiated into replacement cells have been rejected[2], and many clinicians now believe that the cells act by releasing molecules that cause inflammation, with an attendant growth of oxygen-delivering small blood vessels, in the damaged tissue.

The approach has spawned international commercialization of various forms of the therapy, with companies offering treatments for disorders ranging from Parkinson's disease to heart failure. But the effectiveness of such therapies remains unproven.

The *BMJ* study, led by cardiologist Darrel Francis at Imperial College London, examined 133 reports of 49 randomized clinical trials published up to April last year, involving the treatment of patients who had had a heart attack or heart failure. It included all accessible randomized studies, and looked for discrepancies in design, methodology and reporting of results.

Francis's team identified more than 600 discrepancies, including contradictory claims for how patients were randomized, conflicting data in figures and tables, and statistically impossible results. They also found papers listing the same patients as male and female, and patients reported as having died, yet apparently going on to attend tests and report symptoms. The study did not suggest that any error found necessarily affected a trial's conclusions.

A note-in-proof in the paper points out that four of the papers analysed related to influential trials conducted between 2005 and 2010 by cardiologist Bodo-Eckehard Strauer, who is now retired. His work is currently under investigation by public prosecutors after his former employer, the University of Düsseldorf in Germany, found evidence of scientific misconduct.

The note also refers to a trial called SCIPIO involving a different source of stem cells — purported to be specialized cardiac stem cells developed from the patient's aorta — that was recently called into question. Published in *The Lancet* in 2011 and led by Piero Anversa of Harvard University in Cambridge, Massachusetts, SCIPIO showed encouraging results in the use of these cells in patients with heart failure[3]. But Harvard University is now investigating the integrity of some of the data, and *The Lancet* published an unspecified 'Expression of concern' about the paper on 12 April.

The *BMJ* study comes as two major international phase III clinical trials, which are ▶

designed to conclusively test efficacy, have begun recruiting patients. Cardio3 BioSciences, based in Mont-Saint-Guibert, Belgium, is recruiting 480 patients with heart failure in parallel trials of its 'C-CURE' stem-cell therapy — a preparation of specially treated stem cells that are allegedly capable of developing into heart cells. And the European Commission is sponsoring a Europe-wide €5.9-million (US$8.2-million) trial, called BAMI, which tests patient-derived stem cells prepared according to a standardized protocol. It is recruiting 3,000 patients who have recently had a heart attack. The principal investigators of both studies say that the treatment has been shown to be safe and may be effective.

> ***"I have a lot of hope for regenerative medicine, but our results make me fearful."***

However, questions have been raised over an earlier trial of C-CURE. Last June, three months after Francis's study closed, the *Journal of the American College of Cardiology* (*JACC*) published an early-phase trial of C-CURE, which found "signs of benefit in chronic heart failure"[4]. Francis's team analysed it separately and identified dozens of discrepancies similar to those found in the *BMJ* study. He sent details to *JACC*, but claims that the paper's authors did not answer some of his more important concerns: for example, about an apparent change in the study's 'primary endpoint', a trial's main target, and an apparent inconsistency between patient data and the summary of the results.

Co-author Andre Terzic, a cardiologist at the Mayo Clinic in Rochester, Minnesota, denies that Francis's concerns were not addressed and stresses that his group's findings were peer-reviewed. He adds that the decision to drop the initial endpoint — to measure heartbeat strength by monitoring the movement of radioactive tracers through the heart — was made on the advice of the study's steering committee, which said that such efficacy need be assessed only in a phase III trial. The planned phase III trial has now been authorized by the US Food and Drug Administration and the European Medicines Agency, Terzic says.

Another co-author of the *JACC* paper, William Wijns of the Cardiovascular Centre Aalst in Belgium, who is a member of the Cardio3 BioSciences board, told *Nature* that he is "confident in the science supporting the technology and in the C-CURE clinical trial data". A few weeks after the *JACC* publication, Cardio3 BioSciences announced that it had raised €23 million for a phase III trial in a share offering.

BAMI principal investigator Anthony Mathur, of Queen Mary University of London, says that he wants to clarify definitively if there is hope for the treatment. He adds that the trial was built on "a clear signal of efficacy" in some early-phase trials using a standardized protocol that is publicly available.

Christine Mummery, a cardiac-stem-cell researcher at Leiden University Medical Centre in the Netherlands, says that injecting bone-marrow cells causes inflammation and the development of small blood vessels that might limit immediate damage during a subsequent heart attack. "But it is not clear this helps long-term recovery of the heart, and it does not provide a mechanism for improvement in heart failure," she adds.

Even without solid published evidence of efficacy, many companies are offering various commercial mesenchymal-stem-cell therapies to patients with heart disease. For example, the Okyanos Heart Institute in Freeport, the Bahamas, uses mesenchymal stem cells derived from a patient's fat tissue. Howard Walpole, its chief medical officer, was unavailable for comment, but writes on the company's website: "We strongly believe in the science and results we have seen with adult stem cell therapy for coronary artery disease." He adds that many heart patients "do not have the luxury of waiting many years for exhaustive research to be completed".

CardioCell, based in San Diego, California, uses its own standardized proprietary preparation of mesenchymal stem cells rather than a patient's own cells. The company's president and co-founder, Sergey Sikora, says the preparation is based on a method developed at a Moscow research institute in which the stem cells are kept in low oxygen to hone their ability to stimulate the growth of new blood vessels. CardioCell has also licensed the technology to a company called Altaco in Astana, Kazakhstan. Sikora says that CardioCell is currently not offering therapy ouside its own early-phase trials in heart attack and a type of heart failure in the United States, but Altaco has begun a phase III trial for heart attack.

Francis would like to see more evidence that the treatments work before they are exploited. "I have a lot of hope for regenerative medicine, but our results make me fearful," he says. "When the inevitable clinical advantages come, they may be ignored because these 15 years of unreliable data may have damaged credibility." ■

1. Nowbar, A. N. *et al. Br. Med. J.* **348,** g2688 (2014).
2. Laflamme, M. A. & Murry, C. E. *Nature* **473,** 426–335 (2011).
3. Bolli, R. *et al. Lancet* **378,** 1847–1857 (2011).
4. Bartunek, J. *et al. J. Am. Coll. Cardiol.* **61,** 2329–2338 (2013).

SECURITY

US biodefence facilities ramp up

Government effort to produce vaccines on demand raises questions about cost and strategy.

BY SARA REARDON

The future of the US government's biodefence strategy sits in a warehouse in rural Texas. A dozen gleaming-white trailers, each about the length of a bus, hold equipment for producing millions of doses of medical countermeasures against some of the world's deadliest threats. These mobile clean rooms can be configured to manufacture vaccines against pandemic influenza or antidotes to biological, chemical or radioactive agents. Each room can be unplugged from the pipes that supply sterile air and cell-culture media, pushed across the warehouse, and connected to a new production line — ready in days to make a different product.

The US$286-million site at Texas A&M University in Bryan is one of three new biodefence centres created by the US Department of Health and Human Services (HHS). It will start making its first vaccine this summer. Once completed in 2017, it will be able to make 50 million doses of flu vaccine in just four months — capacity that biosecurity experts say the United States needs to prepare for future pandemics.

Yet some worry that the Texas lab and its counterparts form a system that is too disjointed to deliver as promised. Others argue against expanding capacity to produce countermeasures to biological or chemical threats, in part because few effective antidotes exist. "They're going to have a lot of challenges," says Keith Wells, a consultant at BioProcess Technology Consultants in Woburn, Massachusetts.

The $440-million HHS programme, set up in 2012, includes three Centers for Innovation in Advanced Development and Manufacturing (CIADMs): the Texas site; one in Holly Springs, North Carolina, being built by pharmaceutical giant Novartis of Basel, Switzerland; and a facility in Baltimore, Maryland, to be run by biotechnology firm Emergent BioSolutions. Over the next 25 years, the government expects to spend as much as $2 billion on medical countermeasures from the Texas site alone, and up to $23 million per year to

IGOR KRAGULJAC PHOTOGRAPHY

Mobile clean rooms for manufacturing medical antidotes and vaccines at Texas A&M University in Bryan.

stockpile flu vaccines — a cost that could skyrocket if a pandemic strikes.

Gerald Parker, director of the Texas A&M centre, says that the programme's flexible manufacturing is crucial for US biosecurity. "We need to be prepared for all hazards, not just the last one that hit us," he says. Other threats could emerge, such as the coronavirus behind Middle East respiratory syndrome, which surfaced in Saudi Arabia in 2012, or the Ebola virus outbreak currently spreading in West Africa. If one of these viruses sparks a pandemic and a vaccine can be developed, the CIADMs will need to produce it quickly and prepare it for the market.

Such capacity is long overdue, and experts contend that it may be possible to achieve only with government funding. Many large companies have stopped developing vaccines because the public often views them with scepticism, says Robert Kadlec, a former White House director for biodefence policy. And small companies often lack the experience to bring such treatments to market.

Government initiatives such as the CIADMs counter this trend by creating an artificial market for specialized vaccines and antidotes, says Amesh Adalja, a physician at the University of Pittsburgh Medical Center in Pennsylvania who studies biosecurity issues. But incentives are not always effective: on 22 April, Novartis announced the sale of its vaccine programme to drug giant GlaxoSmithKline in London. It also plans to sell its government-funded CIADM in North Carolina.

The US biodefence programme also faces the practical challenge of developing and adopting faster methods to make vaccines. Conventional flu-vaccine production, in which each dose is grown in a chicken egg, is notoriously slow; newer methods that culture vaccines in cells are much faster. In 2012, Novartis' Flucelvax became the first cell-based flu vaccine to gain approval from the US Food and Drug Administration. GlaxoSmithKline and Novartis have developed ways to grow vaccine particles in bird and mammal cells, respectively. And Emergent has partnered with Vaxinnate, a small company in Cranbury, New Jersey, that is working to induce *Escherichia coli* bacteria to make proteins that will raise an immune response against influenza or other agents. "As we move away from eggs, it makes sense not to put our cell cultures in one basket," says Richard Ebright, a molecular biologist at Rutgers University in Piscataway, New Jersey.

STRATEGIC STOCKPILE

The three CIADMs are also working to develop countermeasures against bioterror agents such as smallpox that could be produced on demand. Under the current strategy, "we just hope the bad guys attack us with what we stockpiled", says Brett Giroir, executive vice-president of the Texas A&M Health Science Center, who says that the CIADMs could replenish the existing strategic national stockpile and even expand it over time. The centres' capacity will be put to the test later this year, when the government will place its first orders for medical countermeasures, according to Robin Robinson, director of the Biomedical Advanced Research and Development Authority in Washington DC, which will oversee the centres.

But some question whether these countermeasures, such as advanced smallpox vaccines, are mature enough to be manufactured and stockpiled — or whether they are even needed. The chance of being attacked with a chemical or biological agent, and then being able to use a vaccine to respond to it effectively, is low, says Ebright.

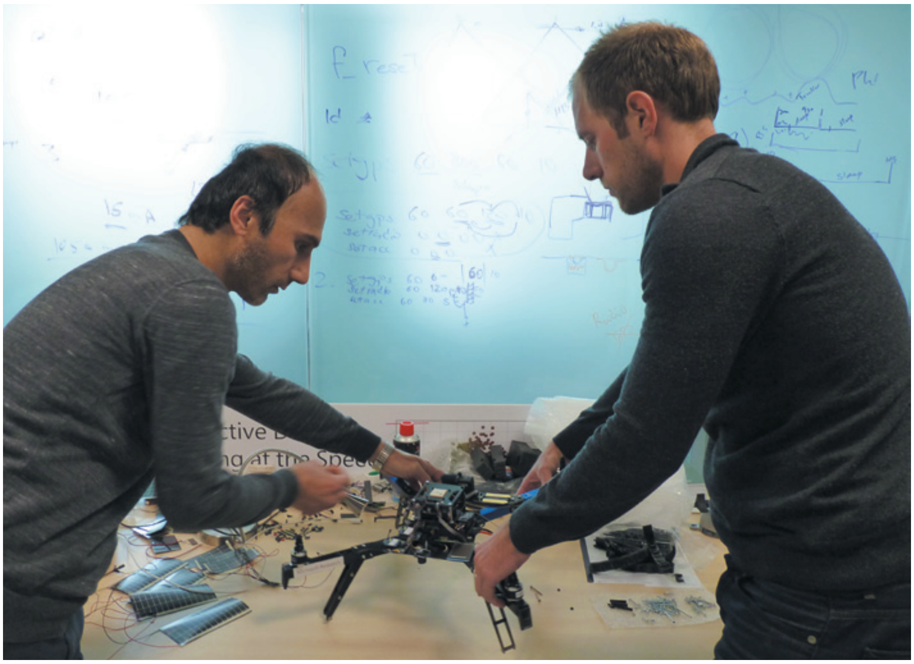
Others say that the United States does not need multiple biodefence centres, pointing to a 2008 report prepared for the Defense Advanced Research Projects Agency, which recommended the creation of a single development and manufacturing facility to serve both the military and civilians. "Rather than one good operation that meets the government's needs, we got three operations that spread the money around," says Philip Russell, a retired major general and former leader of the US Army's medical research command. The HHS centres' supporters say that having three sites creates flexibility and leaves backups in case a site is contaminated or attacked.

Despite concerns about duplicated effort, the US Department of Defense (DoD) began building its own $136-million biodefence facility in 2013 in Alachua, Florida. Once the site is completed in 2015, the military expects to spend more than $20 million on it each year. Kendall Hoyt, a biodefence policy specialist at Dartmouth College in Hanover, New Hampshire, says that the separate effort makes sense: the military requires relatively small amounts of a variety of countermeasures, whereas the HHS needs large amounts of a few products. Moreover, the military needs to respond quickly to the wider range of threats that it faces, says James Petro, acting deputy assistant secretary of defence for chemical and biological defence. Such concerns "led us to recognize that it was important we had a facility that was under DoD control", he adds.

"We need to be prepared for all hazards, not just the last one that hit us."

But Petro says that the separation from the HHS is only temporary. "The intent has been and continues to be that once all the centres are up and running, they'll be operating together as a consortium," he says.

Either way, because they are the first centres of their kind, the HHS facilities have their work cut out for them, says Leonard Cole, a bioterrorism specialist at Rutgers in Newark, New Jersey. A 31 March report from the US Government Accountability Office concluded that it was still too early to tell whether the billions of dollars that stand to be funnelled into the HHS project are a good investment. Nevertheless, "if there's the kind of pay-off we'd hope to come from this, it's worth every nickel", says Cole. ■



MICROSOFT

Scientists at Microsoft are at work on a system that uses aerial drones to track wildlife.

RESEARCH AND DEVELOPMENT

Basic science finds corporate refuge

As many older companies move away from fundamental research, young technology firms are picking up the slack.

BY NADIA DRAKE

Microsoft makes Windows, Word and the Xbox. But it also employs scientists who make sensors to stick on the hides of elephants and rhinoceroses.

The sensors enable aerial drones to track endangered animals and record changes in their movements that might signify threats, says Lucas Joppa, a conservation biologist at Microsoft's research centre in Cambridge, UK. Joppa uses these massive, complex data sets to inform possible conservation solutions, such as systems that warn responders about poaching. He publishes his research freely. "I get to do world-class science, but I also get to interact with one of the world's most successful businesses," he says.

The ranks of scientists like Joppa are growing, according to economists and observers of trends in science and technology. They suggest that a newer generation of technology-oriented companies — many sitting on mountains of cash after years of growth and market success — have enough money to take a chance on the curiosity-driven research of basic science.

"Again and again, basic research has payoffs for applications," says Jennifer Tour Chayes, managing director of Microsoft's research centres in New York and Cambridge, Massachusetts.

Statistics compiled by the US National Science Foundation (NSF) suggest a jump in industrial funding of basic research beginning in 2006 (see 'Corporate masters'). In a climate of stagnant federal and university funding, the increase stands out. Even as some companies have trimmed their research units, others seem to have bolstered them, says Josh Lerner, an economist at Harvard Business School in Boston, Massachusetts. "For every Pfizer cutting basic research, there has been a Google picking up the slack," he says. The trend, if it solidifies, would signal a reversal from the trajectory of the past several decades, which saw industry support for basic research languish. Many cite the dismantling, from 1996 onwards, of Bell Laboratories in Murray Hill, New Jersey — the iconic industrial research centre that invented the laser, the transistor and radio astronomy — as indicative of this larger malaise.

But tech giants such as Microsoft seem to be leading a corporate-research revival. Spread across the globe, the company's research arm comprises roughly 1,100 scientists in fields as varied as ecology, bioinformatics and the social sciences — as well as the computer scientists and mathematicians one might expect. Researchers there operate with few restrictions on inquiry or publication, although they are expected to produce relevant, insightful work — whether for product development or to advance the understanding of nature.

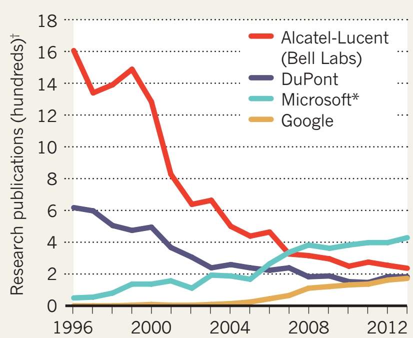
Companies in California's Silicon Valley are following suit. Google, based in Mountain View, brings dozens of outside scientists to its offices for as many as 18 months at a time and spends more than US$30 million annually on grants and fellowships. On 17 April, Twitter, in San Francisco, announced the winners of a programme to let scientists answer research questions using Twitter data. And in December, Facebook, in Menlo Park, tapped Yann LeCun, a computer scientist at New York University, to lead its new artificial-intelligence research group. "There was of a bit of a period until recently where there were very few places in industry where you could do real research," says LeCun.

Another hint at the changing landscape comes from an analysis of company-published research papers. Since 1996, the number of papers associated with older, manufacturing-oriented companies has plummeted, while publications involving younger technology and Internet companies have increased (see 'New kids on the block'). Lerner says the trend is "a changing of the guard".

Still, some observers question whether the jump in funding for basic research is real. The NSF numbers, for example, rely on a survey sent to roughly 40,000 companies that are selected each year from a pool of nearly

NEW KIDS ON THE BLOCK

The number of research publications from some younger technology companies is on the rise, as those from some older companies decline.



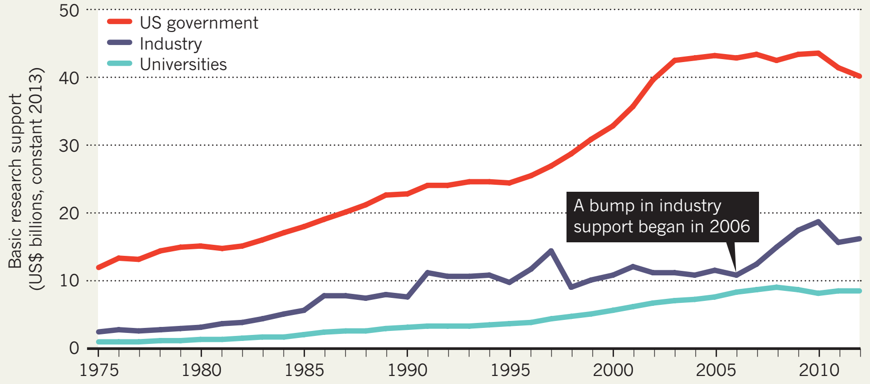
*US and Asia research divisions.
† Publications with at least one company-affiliated author, according to Scopus, a database of more than 20,000 journals.

SOURCE: SCIVAL/ELSEVIER

SOURCE: NSF

CORPORATE MASTERS

As federal support for basic research has flagged in the United States, industry funding has picked up the slack.



2 million. Because the respondents are kept anonymous, it is impossible to know which companies are included in a given year. Companies also self-report how much they invest in basic and applied research — so the upswing in funding could have as much to do with how companies define basic research as anything else, says Matt Hourihan, director of the research and development budget and policy programme at the American Association for the Advancement of Science in Washington DC. "But," he adds, "it could be a reflection of reality."

BACK TO BASICS

The terms 'basic' and 'applied' hark back to Vannevar Bush, a scientist who served as head of the US Office of Scientific Research and Development during the Second World War. "Basic research is performed without thought of practical ends," Bush wrote in *Science, the Endless Frontier*, his 1945 report for the US government. That definition is still pervasive, but companies take slightly different views on how to define and classify projects — Joppa's wildlife-sensor project, for example, fits Microsoft's definition of basic research even though it informs an application (albeit one Microsoft has not historically been interested in). "It's multidisciplinary, disruptive, blue-sky research at its finest," says Chayes.

Some companies, such as Intel, in Santa Clara, California, are explicit about focusing on applied research. At Google, by contrast, scientists are encouraged to pursue "use-inspired basic research", says Alfred Spector, vice-president of research at the company. That means employees often pursue fundamental inquiries that are related to Google's business goals, such as the neuroscience of language-learning or speech perception.

Scientists point to IBM's research lab in Zurich, Switzerland, as an exemplar of innovation through company-funded basic research. A fundamental curiosity about what surfaces look like on an atomic scale led to the 1981 invention of the scanning tunnelling microscope, which in turn helped to usher in the field of nanoscience. From there, the IBM team moved on to build the atomic force microscope, which can image the bonds between single atoms.

Although the Zurich lab is small, four of its members have earned Nobel prizes, and the lab boasts a handful of highly competitive European Research Council grants. Scientists there address basic research questions in physics and biology, as well as slightly more applied topics, such as quantum cryptography, big data and cloud computing. The applied research can sometimes lead back to basic questions, says lab director Matthias Kaiserswerth.

"Basic research, as far as we are concerned, is something where we really want to advance the knowledge of mankind," he says. "What we find is that even if we do applied research, there's always some fundamental question we want to answer." ■

MORE ONLINE

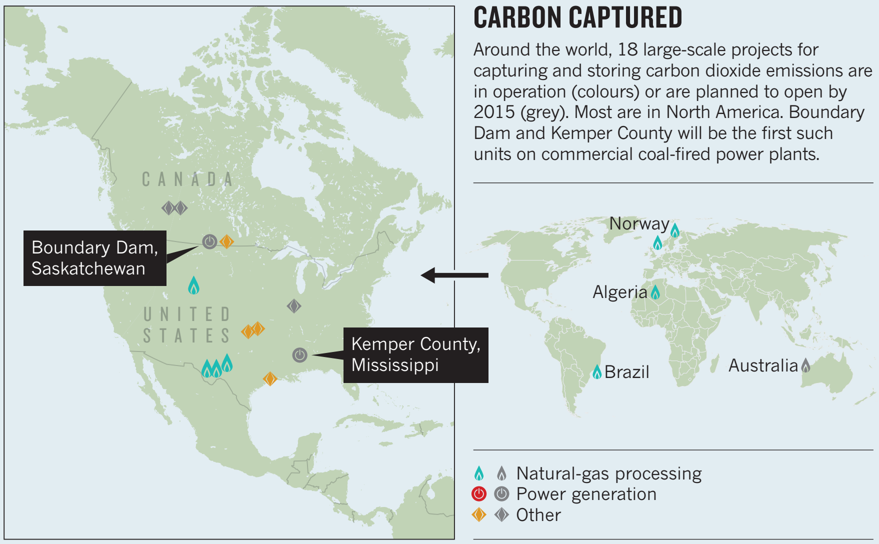
REINHARD DIRSCHERL/ALAMY

VIDEO OF THE WEEK

Sailfish bills nudge as well as slash prey go.nature.com/qrsizh

MORE NEWS

- Clues to sleeping sickness found in genome of tsetse fly go.nature.com/skxe6d
- The tricks used by corals to adapt to warming seas go.nature.com/y4qx19
- Gene therapy helps cochlear implants go.nature.com/iqiiiz



CARBON CAPTURED

Around the world, 18 large-scale projects for capturing and storing carbon dioxide emissions are in operation (colours) or are planned to open by 2015 (grey). Most are in North America. Boundary Dam and Kemper County will be the first such units on commercial coal-fired power plants.

SOURCE: GLOBAL CCS INST.

ENERGY

Two plants to put 'clean coal' to test

Large-scale carbon-capture projects set to go live this year.

BY RICHARD VAN NOORDEN

When Unit 3 at the Boundary Dam Power Station in Saskatchewan, Canada, switches on later this year after a lengthy refit, it will mark a historic moment for dirty coal power. It will be the first time that a commercial-scale plant supplying electricity to the grid captures and stores a large fraction of its carbon dioxide emissions.

"We're getting interest from all around the world," says Robert Watson, chief executive of provincial-government-owned plant operator SaskPower Corporation in Regina. The firm plans to sell about 1 million tonnes of CO_2 a year — up to 90% of Unit 3's emissions — to oil company Cenovus Energy of Calgary, Canada, which will pipe the compressed gas deep underground to flush out stubborn oil reserves.

The project — meant to launch in April but pushed back by renovation delays — looks set to start just ahead of a larger, newly built advanced coal plant in Kemper County, Mississippi, run by Mississippi Power, a subsidiary of Southern Company in Atlanta, Georgia. That plant, which will turn the low-grade coal lignite into burnable gases, is designed to capture 3.5 million tonnes of CO_2 each year, or about two-thirds of what it will produce. Like Boundary Dam, it will sell the gas for oil recovery.

Concerns about the costs of carbon capture and storage (CCS) persist. The technology has progressed extremely slowly, reflecting a lack of enthusiasm for financing large projects. Yet the International Energy Agency (IEA) in Paris says that without CCS it could cost 40% more to decarbonize electricity enough to meet international climate targets by 2050. But the two projects are giving CCS supporters something to cheer about. "This year is really important for CCS," says Stuart Haszeldine, a geologist at the University of Edinburgh, UK.

The concept is hardly science fiction: the oil industry has been piping CO_2 underground to aid in oil recovery for almost half a century. It already purchases some supplies from industrial projects that routinely separate out the gas during procedures such as processing natural gas or making fertilizer (see 'Carbon captured'). And small-scale CCS pilots have shown how solvents can capture CO_2 from power-plant turbines. "The technology is ready to go," says Howard Herzog, a CCS researcher at the Massachusetts Institute of Technology in Cambridge. "The problem is that policies aren't in place to make projects economic."

The extra equipment and energy needed to separate and store CO_2 mean that neither of the pioneering power projects comes cheap. Both have relied on a combination of government subsidies, electricity price hikes and particular local circumstances to get going as other attempts have fallen by the wayside. The Kemper County project is costing US\$5.2 billion, and Mississippi Power expects to increase local electricity prices by 24% to claim back about half its costs. The Boundary Dam refit will cost Can\$1.3 billion (US\$1.2 billion), and SaskPower — the sole electricity supplier in the province — hopes for a 15.5% price increase over the next three years. SaskPower decided to retrofit its plant in part to satisfy Canadian regulations that will cap emissions from new and old units from 2015.

Boundary Dam will also aid research, by separating off a side-stream of CO_2 for an experiment called Aquistore. This project will inject up to 300,000 tonnes of gas per year, depending on Cenovus's operations, into a 3.5-kilometre-deep saline aquifer (a thin layer of brine-soaked porous rock) for storage. Unlike the oil firm, Aquistore will closely monitor what happens to the gas.

In the bigger picture, CCS progress remains stagnant. Coal-burning alone belched out 15 billion tonnes of CO_2 worldwide in 2012, 43% of total CO_2 emissions; thousands of CCS projects would be needed by 2050 to reach targets to put a serious dent in emissions. In 2009, the IEA published a road map calling for 100 large CCS projects by 2020, but last July, with projects failing to materialize, it downgraded that to just 30. Experts say that is still ambitious. The problem is that large emitters such as power stations and steel or cement works have no economic incentive to capture and bury their waste gases. "The slow progress with CCS reflects very closely the level of global climate-policy ambitions," says Juho Lipponen, head of the IEA's CCS unit. For example, there is no appetite for taxing carbon emissions.

Some argue that there is no point in putting high prices on carbon or setting stringent emissions standards while CCS technology is uncertain, expensive and yet to be proven on a large scale. But CCS experts retort that capture technology will get cheaper rapidly as engineering experience accrues, and that it will be possible to store the gas, given the right geological settings.

The two coal-CCS plants, however costly, are globally significant because they could help to end this vicious cycle, says Vivian Scott, who studies CCS policy at the University of Edinburgh. If they are successful, they could demonstrate that CCS can be delivered at commercial scales, he says, and "break the catch-22". ■ SEE EDITORIAL P.7

CORRECTIONS

The opening image in the News Feature 'The plastics puzzle' (*Nature* **508,** 306–308; 2014) wrongly described bisphenol A as a polymer; it is not. The citation for reference 9 in 'Designer reefs' (*Nature* **508,** 444–446; 2014) should have read *Proc. R. Soc. B* **276,** 4139–4148 (2009).

SMALLPOX WATCH

Frozen mummies and envelopes of scabs could contain remnants of one of history's most prolific killers.

BY SARA REARDON

In 2011, while construction workers were digging a foundation at a site in Queens, New York, their equipment struck against something metal. Then a body rolled out of the rubble. Thinking that they might have unearthed the shallow grave of a murder victim, the workers immediately called the New York chief medical examiner's office, and forensic anthropologist Scott Warnasch drove over with a team to check it out.

The body, which had probably been buried in the cemetery of a nearby church, turned out to be a mid-19th-century mummy — of an African-American woman dressed in a nightshirt and socks who had been exceptionally well preserved by her ornate iron coffin. The find struck the forensics team as odd: a black woman in the mid-1800s was unlikely to have been able to afford such a luxurious resting place.

Then the examiners noticed the lesions and raised bumps that covered the corpse. The marks reminded Bradley Adams, New York City's chief of forensic anthropology, of photos he had seen of smallpox victims. The pricey coffin with its airtight seal, the scientists realized, might have been meant not to preserve the body of a wealthy individual but to quarantine an infection. "We took a step back," says Warnasch. The site instantly changed from potential crime scene to potential biohazard, and the city's public-health department called the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, for advice.

CDC officials reassured the New York investigators that the risk of infection was low. But the agency quickly dispatched epidemiologist Andrea McCollum and a team of scientists in protective clothing to autopsy the body and retrieve tissue samples. No one knows how long the smallpox virus can survive in a human corpse, and McCollum's team hoped to recover DNA or even viable virus particles from the woman's body.

This month, the World Health Assembly — the decision-making body of the World Health Organization (WHO) — will meet in Geneva, Switzerland, and decide when to destroy the only known stocks of smallpox virus, held in deep freezes at highly protected laboratories in the United States and Russia. It is a move that has been delayed since the 1980s, and in all likelihood will be put off yet again. But even if the official stocks of virus are destroyed, the chance remains that other batches of the virus could be hidden in a freezer somewhere — or that the pathogen could re-emerge, zombie-like, from a mummified corpse such as the dead woman found in Queens.

Although the risk of such a virus causing a pandemic is low, says McCollum, "it's a very real concern". And an opportunity. In fact, she and a handful of smallpox specialists rather hope that ancient poxviruses, dead or alive, are still out there somewhere. Researchers jump on tips that could turn up specimens reaching back decades or more, and then scour the remains for clues. Practically, such a find could reveal just how long the smallpox virus, known as variola, can

2, NYC OFFICE OF CHIEF MEDICAL EXAMINER; 3, SARA REARDON; 4, JAMES GATHANY/CDC


2


3


4

1. A frozen 300-year-old mummy, from a group found in an ornate coffin in Siberia, yielded DNA from the smallpox virus.
2. Forensic anthropologist Scott Warnasch examines a 160-year-old corpse that was accidentally unearthed in Queens, New York.
3. An 1876 letter containing a scab intended for vaccination turned up in a Virginia museum, setting off a minor scare.
4. Tests of the mailed scab showed that it contained DNA from a virus related to smallpox, but it was harmless.

survive under different conditions. And even if samples do not yield viable virus, researchers could potentially recover DNA from past infections that will add to our picture of how the scourge has evolved, perhaps offering insight into how we might mount defences should smallpox ever re-emerge.

TRACES FROM HISTORY

Smallpox has a reputation as one of the worst diseases in history: it spreads quickly through the human population and kills about one-third of those infected. Although it has afflicted people throughout the world, Native Americans and Africans suffered in extreme ways: some populations all but vanished after contracting the disease from European settlers between the sixteenth and eighteenth centuries. In 1966, when an estimated 10 million to 15 million cases were still occurring worldwide per year, the WHO decided to step up its strategic vaccination and containment campaigns and, by the end of 1977, had eradicated the disease.

Yet remnants of variola remain scattered around the world. Signs of the disease, including skin lesions, virus particles and smallpox DNA, have turned up in stored human scabs and corpses as old as the 3,200-year-old mummy of Rameses V (ref. 1). But no one has ever reported the recovery of live virus from a dead body.

To the CDC's disappointment, the woman from Queens was no exception: the smallpox DNA in her corpse had degraded, leaving no trace of the virus. In theory, other mummies could harbour the virus, although the risk of infection is slight, says Peter Jahrling, a virologist at the US National Institute of Allergy and Infectious Disease in Frederick, Maryland. "If you dig up one of these things, you might want to take some precautions," he says. Cold climates, in particular, may preserve the virus better. "I think it's plausible to imagine that virus might exist in mummies in cold crypts," says Jahrling.

> ***"No one knows how long the virus can survive in a human corpse."***

The virus is extraordinarily stable in human tissue, notes D. A. Henderson at the University of Pittsburgh Center for Health Security in Baltimore, Maryland, who led the WHO's smallpox eradication efforts in the 1960s and 1970s. When the characteristic bumps on a person's skin burst, blood coagulates at the site and fibrous proteins weave a mesh around the virus, trapping it inside a scab. Even after the scabs flake off, the caged virus is unlikely to become airborne or to stick to a person's skin if touched. This stability allowed physicians to develop an early form of immunization known as variolation, in which a small bit of a smallpox scab was inserted into a cut in the skin — causing, for reasons that remain unclear, a milder case of smallpox than a normal infection would have done.

Variolation was used into the twentieth century, especially in poorer communities, although in the nineteenth century many physicians switched to vaccinating with the related, but much less harmful, vaccinia virus. With both viruses, scabs were a useful tool: doctors carried them around in their cases and the scabs could even be mailed, as the Virginia Historical Society (VHS) Museum in Richmond discovered in 2010 when it came across an unusual envelope in its collection. In a letter dated 1876, William Massie, who had recently moved to Richmond, Virginia, wrote to his father Henry in Charlottesville. "[I] hope that this will reach you in faculty of time," he had scrawled in spidery handwriting. Pinned to the letter "so that you cannot lose it as you did before" was a twist of metal foil containing a thumbnail-sized scab that was "perfectly fresh and was taken from an infant's arm yesterday". It should be enough to vaccinate 12 people, the younger Massie wrote.

VHS president Paul Levengood thinks that the artefact is an enchanting glimpse into history. "It's one little story of one family trying to protect itself from something that was this horrifying bane of human existence," he says. "Charlottesville would have been the hinterlands in 1870s; maybe you had access to the vaccine in the big city, but your family wouldn't, so here you were trying to keep them safe."

The museum displayed it along with its letter in an exhibit of "oddities". A few months later, Levengood received a call from the CDC. Someone had alerted the agency after

PATRICK LANDMANN/SPL

The skin of Rameses V is covered with bumps, suggesting that he died of smallpox.

seeing a review of the exhibit in the newspaper.

McCollum and a colleague drove up to collect the scab, hopeful that this one would contain viral DNA. Molecular virologist Inger Damon, who heads poxvirus research at the CDC, says that the oldest known viable sample of variola is from 1939. "To get a sample that came from the 1800s would be enormously important," she says. Studying the virus's evolution could reveal when its ancestor emerged from an animal, or whether poxviruses have evolutionary tricks for stepping up potency, which may be relevant when studying viruses such as monkeypox, which is of increasing concern in Africa.

The museum's scab did contain degraded viral DNA, yet it probably belongs not to variola but to a closely related poxvirus. The CDC group is still waiting for the results of the complete 186,000-base-pair genome sequence "I think they were genuinely disappointed," Levengood says. Just to be safe, the researchers irradiated the scab before returning it to the museum, where it is now kept in a plastic lab tube — albeit a crumble of its former self because of the radiation.

The researchers at the CDC still hope that they will come across a good DNA sample from such a source. "We have little snippets of information from vaccine scabs that fell out of books or were squirrelled away in attics," says Damon. One such cache was stashed inside an envelope in an old book in a New Mexico library, and fell out when a librarian opened the volume in 2003.

Stored scabs have provided the best evidence for how long the virus can live. In an study begun in the 1950s, researchers in the Netherlands collected scabs freshly shed from people with smallpox and kept them in envelopes. They were able to isolate variola virus for 13 years, when they ran out of scabs, and the experiment has not been repeated. Nevertheless, the findings do show that the virus can survive for more than a decade in a temperate climate[2].

A more likely source of infectious virus would be frozen bodies. Influenza viruses seem to be able to survive freezing in lakes and may thereby infect migrating birds[3]. And in February 2014, researchers studying Siberian permafrost reported that they were able to resurrect a giant 30,000-year-old virus that infects amoebae[4]. As global warming continues to melt the permafrost, it is possible that viral diseases could be released from the ice, the authors say.

SIBERIAN MUMMIES

Human activities such as drilling, construction and archaeological digs too, could unearth a viral cache. In 2004, a team of anthropologists hunting for mummies in Yakutia in Eastern Siberia turned up a gravesite containing five well-preserved frozen mummies thrown together in an ornate wooden coffin. The group — probably a family — seemed to have died suddenly in the early 1700s.

The best-preserved body, that of a young woman, had traces of variola DNA in her tissue. The DNA was too degraded for the scientists to reconstruct its entire sequence, but there was enough to show that it was a separate strain from those that circulated widely in Europe and Asia during the twentieth century. The scraps of DNA allowed the researchers to expand our knowledge of the evolutionary history of the smallpox virus, and such studies could eventually help to show how it spread around the world[5].

Scientists at Russia's VECTOR lab in Koltsovo had been thinking about this possibility years earlier. In 1991, a team set out for another village in Yakutia to try to isolate virus DNA from a handful of corpses that had been unearthed by a flood. The researchers were unable to recover any viable virus, or even any dead virus that had retained its shape. "It was a great disappointment," Henderson says. The tissue was well preserved and the virus should have been present. But the Russian scientists were among the best in the business, he says. "If they couldn't get it, it was ungettable."

The Russian researchers say that they have not repeated the expedition. And such hunts for live virus remain a low priority for funding organizations such as the WHO. Of more immediate concern is the ease with which someone could synthesize the virus in a clandestine lab: variola's genome sequence has been available since the 1990s[6]. WHO assistant director-general for health security Keiji Fukuda says that member states will discuss that possibility at the World Health Assembly meeting this month and put together a working group to assess the extent of the threat.

Another concern is that smallpox could escape from a secret cache. Few biosecurity specialists believe that the two stocks kept at the CDC and VECTOR are the only ones in existence. For instance, variola could very well be in the freezer of someone who defected from the Soviet Union, says Jahrling. Perhaps even more worrying is that someone could fairly easily tinker with an existing poxvirus to change its host species, make it more resistant to drugs or coax it to spread more easily.

All those possibilities — along with the chance that ancient smallpox could resurface — has led some researchers to argue that stocks of the virus should stay available. Most people born since the virus was eradicated are not vaccinated against the disease, partly because there are risks associated with the vaccine. And keeping smallpox around to allow the testing of new treatments for the disease makes sense, Jahrling argues. His group has been developing antiviral drugs to treat a smallpox infection after it occurs. "I'm impassioned to keep research going for a while," he says.

Damon adds that newer, safer vaccines are being developed and that having the original virus around could help with testing these as well. She hopes that the decision to destroy US and Russian stocks will once more be delayed. "I don't think we will ever know everything we could possibly know."

In the meantime, researchers continue to piece together the snippets of history they can glean from the occasional corpse or scab. "It's one of these great examples," Levengood says, "of a tiny object that can open up this enormous story." ■

Sara Reardon *covers biomedicine for* Nature *in Washington DC.*

1. McCollum, A. M. *et al. Emerg. Infect. Dis.* **20,** 177–184 (2014).
2. Wolff, H. L. & Croon, J. J. *Bull. World Health Org.* **38,** 492–493 (1968).
3. Zhang, G. *et al. J. Virol.* **80,** 12229–12235 (2006).
4. Legendre, M. *et al. Proc. Natl Acad. Sci. USA* **111,** 4274–4279 (2014).
5. Biagini, P. *et al. N. Engl. J. Med.* **367,** 2057–2059 (2012).
6. Massung, R. F. *et al. Virology* **201,** 215–240 (1994).

COMMENT

BIOETHICS A call to regulate human embryos made for research **p.27**

HISTORY The fitful release of Newton's papers **p.30**

SPACE Cooperation is needed to safeguard the final frontier **p.32**

POLICY Scotland's researchers say that they benefit from being part of the UK **p.33**

SOIL ASSOCIATION

UK farmers in the Duchy Originals Future Farming Programme.

Engage farmers in research

A new wave of small-scale agricultural innovation will boost yields and protect the planet, contend **Tom MacMillan** and **Tim G. Benton**.

Climate change threatens a creaking food system in which harvests are already lagging behind rising demand[1,2]. A sustainable supply of food hinges on agricultural innovation, but current investments neglect a key area for improving yields.

Since the 1970s, agricultural research and development (R&D) has invested mainly in a few research institutes equipped with cutting-edge instruments. For example, the Biotechnology and Biological Sciences Research Council, responsible for much of the public research spending in food security in the United Kingdom, invested 27% of its 2010–11 budget in just three institutes. Multinational seed and agrochemical companies invest billions of dollars to develop products in hopes that they will be used by millions of farmers.

This one-size-fits-all approach has had qualified success. In a 2011 analysis[3], average global crop yields increased by 56% between 1965 and 1985, and by 20% from 1985 to 2005, underpinned by increasing inputs of non-renewable resources.

But advances are slowing. According to a 2013 study[4], yields have plateaued in some of the world's most important food-producing regions, including east Asia (for rice) and northwest Europe (for wheat). In some countries, yields have declined.

The next wave of innovation must be at smaller scales. What one farmer can do to boost yield or efficiency is not necessarily the same as for a farmer hundreds of kilometres away with different soil, microclimate, topology and methods. How well crops and livestock grow depends on the interaction of genes, management and environment. As weather patterns fluctuate, gains in production will depend ever more on innovating in context. Big knowledge flowing from institute to farm must be complemented by local knowledge.

Enhancing farmers' own R&D could reap big rewards for minimal extra cost. Farmers everywhere are practical experimentalists who understand the idiosyncrasies of their land[5]. Modern agronomy evolved out of practices such as rotating crops to rebuild soil nutrients, fertilizing fields with manure, and adding lime to soil to alter pH. Even technologies not invented by farmers — new kit, seeds or chemicals — are adapted by them to fit their circumstances.

Such essential contributions are rarely recognized in official assessments of agricultural R&D. These count farmers as users, rather than makers, of knowledge. When the US Department of Agriculture tots up the US$20 billion that the global private sector invests annually in agricultural R&D, it does not include that done by farmers[6]. Makers ▸

of farm machinery, pesticides, seeds and other 'inputs' invest around 3–11% of their revenue in R&D. Globally, if farmers' innovations were valued at just 0.5% of farming production — worth $4 trillion — that would match formal R&D investment from the private sector.

Some of the best returns can come from helping farmers to assess their own ideas. Until now, such initiatives have been at arm's length from formal science, and almost exclusively in the developing world. Our involvement in a farmer-focused innovation programme in the United Kingdom has convinced us that such participatory R&D could also boost agricultural innovation in rich countries.

GRASS-ROOTS RESEARCH

Farmer-centred approaches are not new. In villages in Kenya, rice fields in Indonesia and other places out of reach from industrialized agriculture, group learning programmes recognize and support farmers as innovators.

The best known of these is the farmer field school approach, in which groups of farmers meet regularly to learn alongside their neighbours. The UN Food and Agricultural Organization set up the first such school in Indonesia in 1989, aiming to reduce rice farmers' reliance on pesticides by enabling them to observe, identify and actively manage pests' natural enemies.

Since then, at least 10 million smallholder farmers have taken part in field schools across Asia, Africa and Latin America. This year, a meta-analysis[7] of 71 projects found that farmers' experiences of these schools vary enormously, with targeted initiatives being more successful than large-scale national programmes. In targeted initiatives, participants gained knowledge, changed practices and consequently netted higher yields and incomes.

Inspired by the approach, a UK programme adapts participatory learning to suit farmers in the industrialized world, who, in many cases, are not short of capital, training or access to knowledge. Piloted in 2012, the Duchy Originals Future Farming Programme is funded by the Prince of Wales's Charitable Foundation, in turn funded by sales of products through the supermarket chain Waitrose. The work is led by two charities, the Soil Association in Bristol (at which T.M. works, and T.B. serves on the programme steering group) and the Organic Research Centre in Newbury. The aim is to help farmers to sharpen their skills as innovators so that they can be more productive with fewer non-renewable inputs — good for the environment and their bottom line.

Groups of 5–15 farmers tackle a problem put forward by a participant and test solutions over up to a year in as many as four workshops on one of their own farms. A facilitator helps to keep meetings on track, and a relevant researcher — crop or animal scientist, agronomist or ecologist — is on hand to advise on experimental design and point out existing studies to avoid redundant work.

So far, 450 UK farmers have piloted 'field labs' for about 20 topics, with results documented publicly online. Their farms range from under a hectare to more than a thousand. Field labs have tested ways to control black grass (a persistent weed that resists herbicides), assessed the economics of keeping hens alive to lay eggs for a second season, and evaluated ways to reduce use of drugs that control liver fluke in sheep.

These field labs do not always provide clear answers because of their small samples and short timescales. Field labs raise scientific standards nonetheless: early evaluation suggests that most farmers who have taken part in field labs are eager to engage with formal research. And some have yielded useful lessons. In one, vegetable growers tested composts that do not include peat (which releases greenhouse gases when mined). In contrast to prevailing views, the farmers deemed these peat-free composts commercially viable. And the agronomists learned how labour savings from easily handled materials can outweigh business benefits of higher germination rates.

Grass-roots R&D projects are cropping up elsewhere in the developed world. A participatory breeding programme sponsored by the European Union (EU) has recruited farmers to develop crops that can be grown more sustainably. Organic and small-scale farmers in Italy and France (and in some developing countries) are testing and selecting varieties of barley, beans, broccoli, maize (corn), tomato and wheat.

Animal scientists in Denmark adapted the farmer field school approach to develop 'stable schools'. Four groups of around five farms each worked together to assess changes to herds' housing, hygiene and milking practices and reduced use of antibiotics[8]. A study funded by the European Commission is evaluating 17 'learning and innovation networks' for sustainable agriculture.

Apps, software and websites that recognize farmers as innovators, not just managers, are also on the rise. In the United States, FarmHack.net is an open-source community in which mainly small-scale farmers share know-how, tools and designs. Recent posts include advice on affordable aerial imaging and guides for repurposing old equipment.

More lessons are coming from the developing world. CABI, an intergovernmental agency, is training community 'plant doctors' who help farmers to identify pests and diseases and to enter the information in open-access databases that could be used to control pests or track epidemics.

Research funders are waking up to the advantages of asking farmers what they need to know. In the United Kingdom, the main farming bodies convened a consultation called Feeding the Future that identified topics such as precision agriculture and animal-disease management as practical priorities[9]. But we believe that field labs could boost farmers' productivity by supporting low-cost innovations that fly below the radars of large research institutions. When farmers produce knowledge, they are more likely to adopt new practices, and their insights are more likely to be relevant to local conditions.

TESTING GROUND

Field labs attract innovative farmers — early adopters who can spread best practices. The challenge now is to evaluate and popularize the approach. In Europe, at least, the

MIKKEL OSTERGAARD/PANOS

Participants in a farmer field school in Uganda learn about manure.

PASCAL GOETGHELUCK/SPL

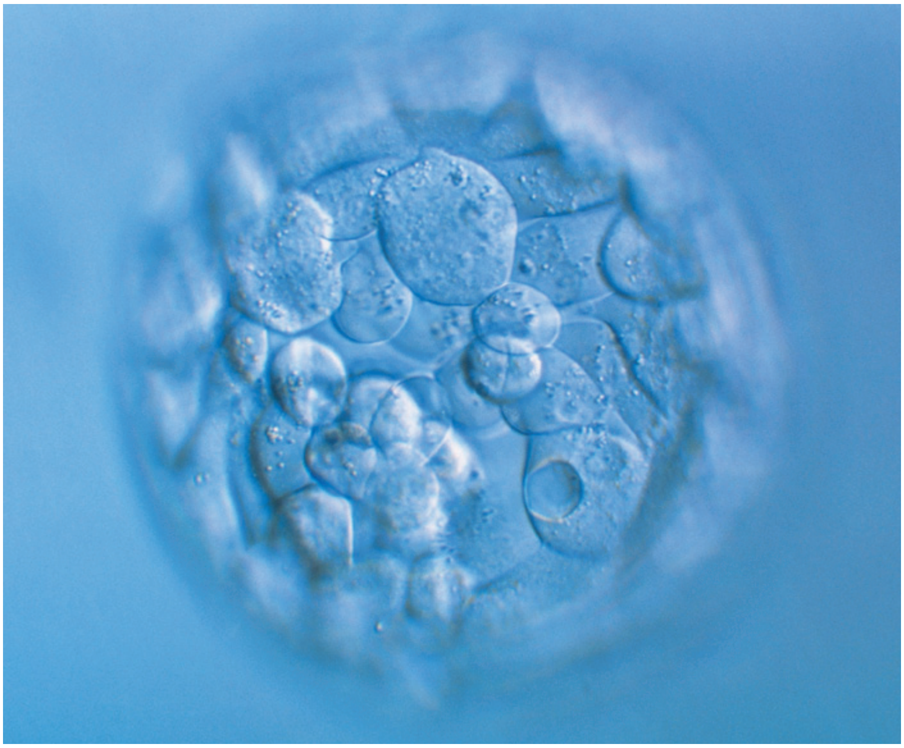
moment may have arrived. Linked to the latest round of reforms to the Common Agricultural Policy and the Horizon 2020 research programme, the EU launched the European Innovation Partnership for agricultural productivity and sustainability. It aims to promote bottom-up approaches by linking farmers, researchers, businesses and other stakeholders into groups charged with finding solutions to shared problems. With billions of euros earmarked for food and farming research over the next six years, the impact of this initiative is potentially enormous.

The European Commission has set out principles for this approach. Whether it flies or fails depends how EU member states rise to the challenge. For this initiative to succeed, governments must opt to spend a proportion of their rural development funds on supporting grass-roots training and learning by actual farmers — beyond the established partnerships with farmers' suppliers, customers and political representatives. Governments should back brokerage services that help farmers to team up with relevant researchers on their own terms, and enable them to navigate the maze of bureaucracy that will probably stand between them and this invaluable seed investment.

The time has come to decentralize, diversify, and enrich agricultural R&D. Farmers — not scientists, outreach workers or salespeople — are the essential players in any agricultural innovation system. Helping them will put food on the world's tables. ■

Tom MacMillan *is director of innovation at the Soil Association, Bristol, UK.*
Tim G. Benton *leads the United Kingdom's Global Food Security programme and is professor of population ecology at the University of Leeds, UK.*
e-mails: tmacmillan@soilassociation.org; tim.benton@foodsecurity.ac.uk

1. Challinor, A. J. *et al. Nature Clim. Change* **4,** 287–291 (2014).
2. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. *PLoS ONE* **8,** e66428 (2013).
3. Foley, J. A. *et al. Nature* **478,** 337–342 (2011).
4. Grassini, P., Eskride, K. M. & Cassman, K. G. *Nature Commun.* **4,** 2918 (2013).
5. Standing Committee on Agricultural Research. *Agricultural Knowledge and Innovation Systems in Transition* (European Commission, 2012).
6. Fuglie, K. O. *et al. Research Investments and Market Structure in the Food Processing, Agricultural Input, and Biofuel Industries Worldwide* (US Department of Agriculture, 2011).
7. Waddington, H. & White, H. *Farmer Field Schools: From Agricultural Extension to Adult Education* 17 (International Initiative for Impact Evaluation, 2014).
8. Vaarst, M. *et al. J. Dairy Sci.* **90,** 2543–2554 (2007).
9. Pollock, C. *Feeding the Future* (Joint Commissioning Group, 2013).



A human blastocyst, an early-stage embryo of the kind used in research, comprising 150 or so cells.

Regulate embryos made for research

As technical barriers fall, the United States should adapt existing measures to govern the generation of human embryos for research, says **Insoo Hyun**.

Three independent research teams have now used cloning technology to make human embryonic stem cells carrying the genomes of existing people. The first announcement, using genomes from fetal and infant cells, came last year[1]. The next two reports have emerged in the past month, detailing human embryonic stem cells that were custom made from cell samples derived from living adults, including a 75-year-old man[2] and a 32-year-old woman with type 1 diabetes[3].

This repeated cloning of embryos and generation of stem cells, now using cells collected from adults, increases the likelihood that human embryos will be produced to generate therapy for a specific individual. The creation of more human embryos for scientific experiments is certain. Regulatory structures must be in place to oversee it.

These accomplishments were made possible by numerous tweaks, and by mastery of difficult techniques and of the administrative work required to collect enough eggs from healthy young women. Each research team inserted nuclei taken from human skin cells into unfertilized eggs from which the original nuclei had been removed. These constructions grew in a dish into early-stage embryos — hollow balls of about 150 cells. Dozens of embryos and about ten stem-cell lines were derived across the three studies.

With reliable techniques now available, researchers will want to compare the therapeutic potential of cloned embryonic stem cells with stem cells made from the easier technique of reprogramming adult cells. Doing so could require the production of cloned embryos from people with a range of diseases. Experiments that depend on custom-made embryos could also be used to investigate complex human diseases, infertility and, perhaps eventually, to generate genetically matched replacement ▶

tissues for people who lose organs to disease or degeneration.

For some, this raises two dangerous spectres: cloned human babies, and a future in which human embryos are callously created and destroyed for various kinds of research. Both scenarios can be avoided. Current policies (and probable biological barriers) are sufficient to mitigate the first. The second can be headed off by adding a provision to existing oversight structures.

NO CLONED PEOPLE

Earlier variants of the embryo-cloning technique, called somatic-cell nuclear transfer (SCNT), produced domestic animals including cows, dogs and, most famously, Dolly the sheep. Efforts in humans have focused on cloning embryos not for pregnancies (reproductive cloning) but for laboratory work (research cloning). It is important to keep these two purposes separate.

Human reproductive cloning is already preemptively illegal in more than 30 countries and in 13 US states. Even in US states where reproductive cloning is not banned, it would require regulatory oversight by the Food and Drug Administration (FDA). The frequently observed ill health of cloned mice, sheep and other animals make safety concerns alone sufficient for the FDA and similar agencies in other countries to block attempts to clone a human baby. This is the position of scientific guidelines issued by the International Society for Stem Cell Research, the US National Academy of Sciences and the 1997 National Bioethics Advisory Commission.

Debate is raging, however, over a related procedure for reproduction: transferring maternal chromosomes from a faulty egg into a healthy egg from which the chromosomes have been removed, which is then fertilized with a partner's sperm[4]. *In vitro* research in human gametes and successful reproduction in macaques suggest that women carrying mitochondrial diseases could bear genetically related children in this way without passing on defective mitochondria.

Although produced with similar techniques to cloning, the resulting embryo would not be a clone. The radical manipulations of the donor egg and transferred chromosomes, however, might compromise the health of any resulting child[5]. Ongoing animal studies are unlikely to provide sufficient safety data for regulators to decide whether human trials should proceed. Human embryos created for research by nuclear transfer might aid preclinical assessment. In other words, embryos created for research could be used in work that helps women with defective eggs to have children.

Laws governing SCNT work in humans generally reflect the desire to stop human reproductive cloning while supporting valuable science. However, the ethics of creating and destroying human embryos for research are complex. Laboratory-produced human embryos have, at minimum, a symbolic value for most people and should not be used carelessly[6]. Whether the latest advances could lead to research that trivializes embryos is uncertain. Nevertheless, wise policy would allay social fears while maintaining public support for scientific research.

Some countries, notably Germany and Canada, have made it a criminal offence to create embryos for research by any means, including by *in vitro* fertilization (IVF), SCNT or from unfertilized eggs with original nuclei. Australia, Singapore and the United Kingdom allow the creation of research embryos by SCNT, but only under regulation by government agencies. The UK Human Fertilisation and Embryology Authority (HFEA) and Australia's Embryo Research Licensing Committee review and approve proposals to license the creation of embryos for specific scientific experiments. Although scientists complain about delays, these bodies have allowed research to proceed with public confidence, and facilitated crucial advances in assisted reproductive technologies.

> ***"Barriers to obtaining eggs will limit the production of embryos."***

Policies across the United States run the gamut. Polls conducted during the presidency of George W. Bush (which banned the use of federal funds for research on embryonic stem cells created after 9 August 2001) showed that most Americans approved of research on stem cells derived from surplus embryos in fertility clinics but objected to creating embryos specifically for research[7]. Now that President Barack Obama has lifted the funding restriction, today's policy mirrors this ambivalence: public funds may be used only on embryonic stem cells derived from donated surplus embryos originally made for reproduction.

EVOLVING RULES

In 1994, a panel of the US National Institutes of Health (NIH) drafted guidelines for the creation of research embryos but political pressures prevented them from going into effect. These stipulated that creating embryos for research requires, for instance, that: surplus IVF embryos would be unsuitable for meeting scientific goals; research goals would greatly benefit science and society; the minimum number of embryos would be created to answer the question; and their use would be limited to the shortest time necessary, at most within 14 consecutive days of development *in vitro*[8]. These are precisely the types of consideration that national-licensing bodies such as the HFEA take into account, but the United States currently has no such review mechanisms for proposals that seek to create research embryos.

Here, a legacy from the Bush era might help. In the absence of NIH funds (and concomitant ethical guidelines) the US National Academy of Sciences recommended that every organization performing research on embryonic stem cells establish a dedicated review committee and require scientists to submit relevant protocols for its approval before proceeding with experiments. Academic institutes and even for-profit companies voluntarily did so. These committees, which are still in place, seem well suited to assess the creation of research embryos.

Their duties, however, will need to be expanded. Currently, these committees are responsible for tracking provenance and research use of stem-cell lines derived from IVF embryos. In my view, the committees should be required to evaluate scientists' rationales in proposals to create embryos for research. It is possible that different institutions, like different countries, will come to different conclusions about which experiments should proceed, confusing the public or perhaps prompting productive debate.

This approach does have limitations. Stem-cell-research oversight committees might lack the expertise or remit to consider all situations in which embryos might be created, such as reproductive uses of heavily manipulated or cultivated gametes, as would be used for transferring maternal chromosomes from an unhealthy egg. So alternative approaches are also worth considering. Still, review mechanisms work best when they build on existing regulatory practices. Adapting structures already in place will address ethical issues adequately without hindering science unduly.

It is easy to see how the three most recent cloning studies broke scientific ground, but that assessment will become thornier as work progresses. Barriers to obtaining eggs will limit the production of embryos for less-than-cutting-edge research. But it is important to review specifically whether the questions that a custom-made embryo could help to answer justify its creation. ■

Insoo Hyun *is associate professor of bioethics at Case Western Reserve University School of Medicine in Cleveland, Ohio, USA.*
e-mail: insoo.hyun@case.edu

1. Tachibana, M. *et al. Cell* **153,** 1228–1238 (2013).
2. Chung, Y. G. *et al. Cell Stem Cell* http://dx.doi.org/10.1016/j.stem.2014.03.015 (2014).
3. Yamada, M. *et. al. Nature* http://dx.doi.org/10.1038/nature13287 (2014).
4. Vogel, G. *Science* **343,** 827–828 (2014).
5. Nuffield Council on Bioethics. *Novel Techniques for the Prevention of Mitochondrial DNA Disorders: An Ethical Review* (Nuffield Council on Bioethics, 2012).
6. Dworkin, R. *Life's Dominion* (Vintage, 1993).
7. Nisbet, M. C. *Public Opin. Q.* **68,** 131–154 (2004).
8. National Institutes of Health. *Report of the Human Embryo Research Panel* (NIH, 1994).

Isaac Newton, painted by Godfrey Kneller in 1702.

CORBIS

SCIENCE BIOGRAPHY

A voyage round Newton

Mordechai Feingold savours a study on how the fitful release of the scientist's papers shaped his reputation.

The last major private repository of manuscripts and correspondence by Isaac Newton and his circle was held in the extraordinary library at Shirburn Castle in Oxfordshire, UK, seat of the earls of Macclesfield. In the late 1970s, I sought access to the documents while pursuing a doctorate at the University of Oxford. I was politely, but firmly, refused. My mentor, Hugh Trevor-Roper, pleaded with the earl, who countered that the last time an Oxonian had entered the library, a book had gone missing. Trevor-Roper apologized on the university's behalf, and asked when the incident had occurred. "In 1747," came the curt reply.

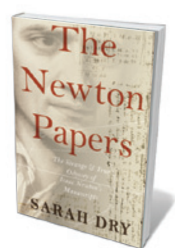
This anecdote illustrates a key element in *The Newton Papers*, Sarah Dry's engaging narrative of the fortunes of the towering mathematician's *Nachlass* — his private papers. These were inaccessible to researchers for more than two centuries after his death in 1727. Had the politician John Conduitt (1688–1737) completed his biography of Newton — based on papers he had inherited through marriage to the scientist's niece — public perception of the great man's persona might well have been different during the eighteenth and nineteenth centuries. In the event, people were left to conjure for themselves their perceptions of Newton's character and extra-scientific interests. Mesmerized by his scientific achievement, Enlightenment savants divinized Newton, and presumed that his character corresponded to his genius.

Access to most of the papers has now been free for four decades, and the Shirburn cache was finally sold in 2000, for well over £6 million (US$10 million), to the Cambridge University Library, UK. Since then, historians have been able to develop a vastly more comprehensive and nuanced understanding of Newton. They have contextualized his scientific pursuits and his life-long interest in areas outside that realm: alchemy, biblical prophecies, and Church and universal history.

Dry's book covers in loving detail and with verve the various permutations of public perception. The speculation began in earnest with the posthumous publication of two of Newton's secret works. In *The Chronology of Ancient Kingdoms Amended* (1728), Newton sought to synchronize sacred and profane history — as reckoned at the time — by shortening Egyptian history by about 1,200 years and Greek history by 500. *Observation Upon the Prophecies of Daniel, and the Apocalypse of St. John* (1733) shows Newton attempting to rationalize the visions of the biblical books of Daniel and Revelation into a coherent prophetic structure, capable of accounting for past, and potentially future, events. Around this time, anecdotal information on Newton's denial of the Trinity also circulated. But neither the content of his posthumous publications nor hearsay about his heterodox religious opinions tarnished Newton's reputation during the Enlightenment.

Starting with Voltaire's relentless campaign to popularize Newton in the 1730s and 1740s, there emerged a view of the mathematician as the greatest and rarest genius, whose character was equal to his mental prowess. In the 1770s, an opportunity to present a more rounded view presented itself when Samuel Horsley, British mathematician and future bishop of Rochester, was allowed access to some of Newton's private papers for inclusion in a new edition of his work. Ultimately, for reasons that are still not clear, none was included.

In the early nineteenth century, the grandeur of Newton's science remained intact. His reputation, however, was tainted by information suggesting that he had had a nervous breakdown, gleaned from the papers of some of his contemporaries, including Dutch astronomer Christiaan Huygens and English philosopher John Locke. The first provocateur was French mathematician Jean-Baptiste Biot. In 1822 he published a short biography of Newton that contained



The Newton Papers: The Strange and True Odyssey of Isaac Newton's Manuscripts
SARAH DRY
Oxford University Press: 2014

Some of Newton's papers, auctioned in 2000.

STEFAN ROUSSEAU/PA

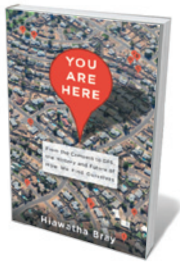
the shocking allegation that Newton had gone mad in the early 1690s, and never entirely recovered — which, Biot insinuated, made him fervently devout and inspired his religious writings. In 1835, an edition of letters and memoirs by the first British astronomer royal, John Flamsteed, brought to light Newton's seemingly heartless treatment of Flamsteed. The revelations engendered debates in England and across the Channel, prompting Scottish physicist David Brewster to spring to Newton's defence. His work culminated in 1855 with a magisterial two-volume biography, which mostly glossed over Newton's heterodox religious views.

Along with her narrative of these debates and discussion of how the history of science evolved in England, Dry offers lively portraits of those who enabled the recovery of the "true" Newton. These include the two earls of Portsmouth who owned his papers — Isaac Newton Wallop, who donated the scientific and mathematical manuscripts to Cambridge in 1872, and Gerard Wallop, who put the rest up for auction in 1936 to pay for death duties and his own divorce. Also discussed are the dealers involved in the dispersal of the papers, and the passionate collectors who vied for a share in the spoils — particularly the economist John Maynard Keynes and the polymath Abraham Yahuda. These two acquired a considerable portion of Newton's alchemical and theological manuscripts in the 1936 sale. Keynes and Yahuda's wife later donated their respective collections to King's College, Cambridge, and the National Library of Israel in Jerusalem, finally allowing free access to Newton's nonscientific papers for the first time.

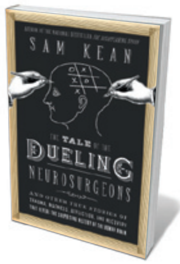
Dry is to be congratulated for furnishing us with a fresh and readable chronicle of the tortuous route that Newton's manuscript took to being made public — ostensibly in accordance with the wishes of the great man. ■

Mordechai Feingold *is professor of history at the California Institute of Technology in Pasadena. His latest book is* Newton and the Origin of Civilization *(with Jed Buchwald).*
e-mail: feingold@hss.caltech.edu

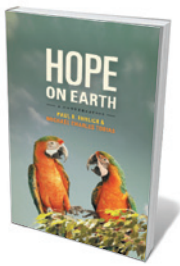
Books in brief



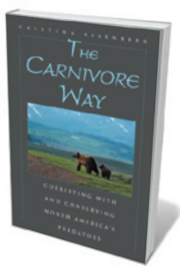
You Are Here: From the Compass to GPS, the History and Future of How We Find Ourselves
Hiawatha Bray Basic Books *(2014)*
Getting lost — that adjunct to exploration — is rare in our minutely mapped world. Hiawatha Bray traces the technologies that have driven us to this state of hyper-visibility. Kicking off with Tudor physician William Gilbert's finding that Earth is a colossal magnet, Bray covers radio and gyroscope, global positioning satellites, Wi-Fi, smartphones, Google maps, chip-and-reader tracking and more. The implications of this Orwellian "locational transparency" are amply discussed, to unsettling effect.



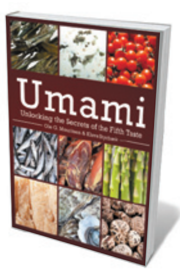
The Tale of the Dueling Neurosurgeons
Sam Kean Little, Brown *(2014)*
Oliver Sacks's 1985 *The Man Who Mistook His Wife for a Hat* (Summit Books) set off a tsunami of interest in brain anomalies and linked behaviours. Sam Kean braves that wave, revealing how intercranial calamities have built neuroscience case by puzzled-out case, gross anatomy to consciousness. However pop the science, there is much to compel. Victorian explorer James Holman, for instance, was blinded by an infection yet, Kean writes, travelled 400,000 kilometres by echolocation (sounding out surfaces by tapping them with a cane) and was cited by Darwin for work on island seed dispersal.



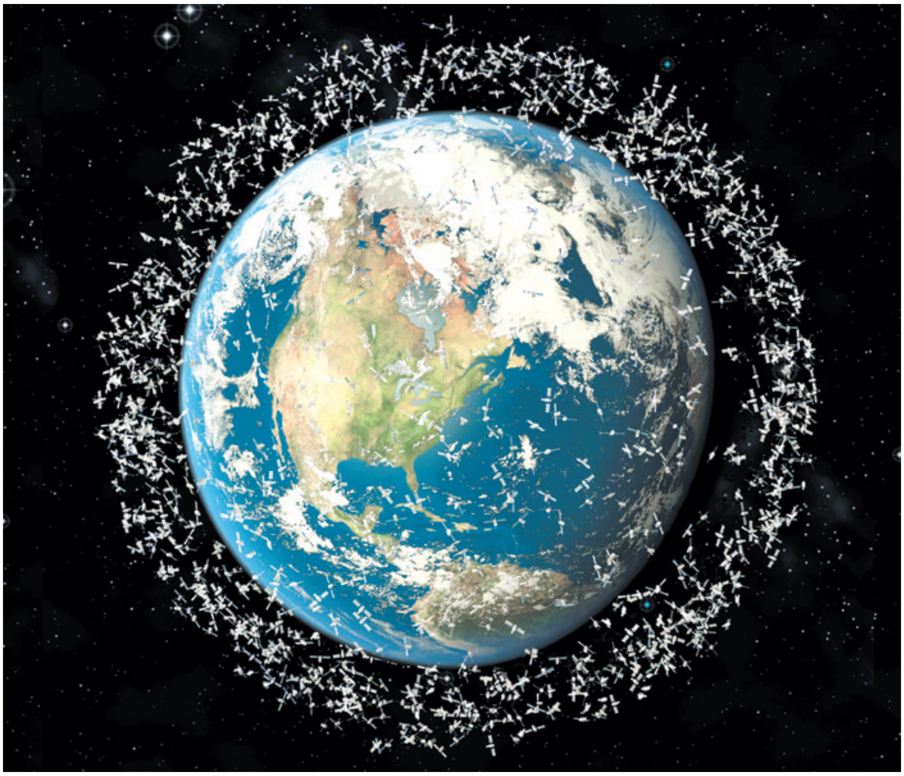
Hope on Earth: A Conversation
Paul R. Ehrlich and Michael Charles Tobias University of Chicago Press *(2014)*
For those wearied by the malaise enveloping high-level negotiations on planetary ills, these 'bilateral talks' between biologist Paul Ehrlich and ecologist Michael Charles Tobias offer relief. Their often heated dialogue on the scientific, ethical and conceptual approaches to our global predicament is cogent on details such as the realities of unsustainable livestock farming, or the broken equation of population and resources. Despite the grim litany, their action plan for humanity — such as limiting family size — is convincing.



The Carnivore Way: Coexisting with and Conserving North America's Predators
Cristina Eisenberg Island Press *(2014)*
The "carnivore way" is a vast, fragmented wildlife corridor stretching along the Rocky Mountains from Alaska to Mexico. In this call for a unified vision in conservation, ecologist Cristina Eisenberg argues that big carnivores such as grizzly bears underpin the corridor's ecological health, and need it in turn for dispersal into new territory. She interweaves multiple skeins of science — on predator population resilience, the success of highway wildlife crossings and more — to build a putative scenario of human–carnivore coexistence.



Umami: Unlocking the Secrets of the Fifth Taste
Ole G. Mouritsen and Klavs Styrbæk Columbia University Press *(2014)*
From Parmesan's savoury tang to the deep complexities of the Japanese soup stock dashi, umami adds a fifth dimension to the flavour universe. Biophysicist Ole Mouritsen, whose *Seaweeds* (University of Chicago Press, 2013) seamlessly meshes science and gastronomy, here reprises that recipe with chef Klavs Styrbæk. Research into areas such as umami's role in appetite regulation alternates with a stunning menu featuring, for instance, braised monkfish liver with peanuts and raspberries. **Barbara Kiser**



ROGER HARRIS/SPL

Satellite deployment must be regulated to keep space usable and safe.

SPACE POLICY

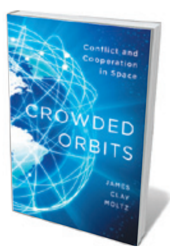
A clearer final frontier

David Southwood finds that a study on safeguarding space is shorter on cooperation than conflict.

As I read policy expert James Clay Moltz's *Crowded Orbits*, an old saying came to mind: 'If all you have is a hammer, you see every problem as a nail.' The proverb points to the risks in knowing a lot from a particular point of view. Having worked on putative and actual space alliances involving India, China, Canada, Japan, Russia, the United States and nearly every European country, I knew a lot; but my limited perspective became clear as I read. So did limitations in the book itself. International space activities — conflictual, competitive or cooperative — look different depending on where you stand.

Moltz's background is military; the focus of *Crowded Orbits* is the international regulation of space access and use, and the avoidance of war in space. Moltz argues for serious geopolitical engagement to protect space against catastrophic threats: actions, intentional or otherwise, that could render regions of space unusable as a result of debris or, less likely, radiation. His understanding of the threats from debris makes the book valuable to anyone involved in space, from astronomer to industrialist to civil servant. Nevertheless, the context given is only part of the story of why the final frontier needs to be kept clear.

Increasingly, space is crucial in civilian infrastructure and terrestrial military activity: communications, geospatial information, environmental monitoring and navigation depend on it. Should not nations cooperate to prevent a debris catastrophe? The United Nations has committees on disarmament and on peaceful uses of outer space, yet neither looks capable of managing this task. At the root of the problem is the dialectic of self-interest and competition versus mutually supportive cooperation. Moltz makes the case that means need to be found.



Crowded Orbits: Conflict and Cooperation in Space
JAMES CLAY MOLTZ
Columbia University Press: 2014.

He examines models of cooperation and concentrates on the International Space Station. His examination of other astronomical and exploration cooperation efforts is somewhat cursory, and there are small errors. The Rosetta mission of the European Space Agency (ESA), for instance, is not a solar mission, but instead will rendezvous with and observe a comet. The lunar probe Chandrayaan-1 was a collaboration between India, ESA, NASA and Bulgaria; and unmentioned is the ground-breaking Double Star Sino-European joint mission.

Failures in cooperation offer many lessons, and Moltz looks at the International Solar Polar Mission almost 25 years ago, as well as the United States' 2012 pull-out from collaboration with ESA on Mars exploration. However, he misses the pivotal case of the International Gamma-Ray Laboratory (INTEGRAL), launched in 2002, the absence of expected levels of US support for which led to the first close ESA Science Programme cooperation with Russia. The distinction between cooperation and competition had, however, emerged decades before, with US refusal to launch the Franco-German communications satellite Symphonie. Europe learnt that it had to go it alone in using space for commercial or near-to-market purposes.

Moltz accepts that space is important to commerce, but readers might query whether he sees how this may drive regulation. The European Union, the largest economic grouping in the world, is this year launching multiple spacecraft; these include members of the Sentinel family of Earth-monitoring satellites. The move is economic, and the use of space is evolving and increasingly focused on societal benefits such as navigation and environmental monitoring. Moltz integrates this less well into his vision of the future.

These quibbles do not detract from Moltz's case for internationally regulating and safeguarding use of space. The way to achieve it, however, could be different from joint interest in exploration or military stand-off. Lessons can be drawn from civil cooperation in areas such as Earth observation. Every nation has an interest in managing our planet better, which requires space data. Is an alternative way of raising consciousness something like the Group on Earth Observations (GEO)?

GEO is Switzerland-based, US-proposed and G8-initiated; through it, 90 countries have signed up to free exchange of space data for Earth monitoring. GEO will never launch satellites, but could its reach lead to a worldwide sense of the need to preserve access to and use of space? As good as *Crowded Orbits* is, it is clear that the future will need a complete tool kit — not just hammers. ■

David Southwood, *a senior research investigator at Imperial College London is chairman of the board of trustees of the London Institute for Space Policy and Law.*
e-mail: d.southwood@imperial.ac.uk

Correspondence

Rise in retractions is a signal of integrity

A stigma should not be attached to the retraction of a scientific paper, as you explain (*Nature* **507,** 389–391; 2014). It should also be emphasized that the rise in retractions over the past few years does not signify a surge in misconduct: on the contrary, it reflects a growing scientific integrity.

Too many academics and journalists conflate retractions with the falsification of results. However, retractions account for less than 0.02% of publications annually — a fraction of the 2% of scientists who admit in anonymous surveys to having manipulated data at least once (see D. Fanelli *PLoS ONE* **4,** e5738; 2009).

The majority of formal retractions have been issued in recent years, with none before the 1970s. A growing number of journals are now prepared to publish retractions, and the apparent increase in retraction rate disappears after correcting for this factor (see D. Fanelli *PLoS Med.* **10,** e1001563; 2013).

Retractions are therefore more logically and usefully interpreted as evidence for the commitment of editors and scientists to remove invalid results from the literature.

Daniele Fanelli *University of Montreal, Canada.*
email@danielefanelli.com

Practical costs of data sharing

Aside from the ethics and etiquette of fully open data-sharing (*Nature* **507,** 140; 2014), there are practical issues that journals still need to address.

One is the cost of sharing data. Both the Public Library of Science and the UK Royal Society recommend the storage repository Dryad, which currently charges US$15 for the first gigabyte of data over its 10-gigabyte limit, and $10 per gigabyte thereafter. However, studies in areas such as neuroscience can generate terabytes of raw data (1 terabyte is 1,000 gigabytes) — a quantity that few labs could afford to upload.

And, given that searching Dryad for 'neuroscience' yields just three papers but 2,286 for 'ecology', a 'one-size-fits-all' data-sharing policy may not work across all disciplines.

Another concern is the availability of new computer code. Researchers often write their own data-analysis code for each new study, but do not always document it fully. Making code usable by others may therefore require considerable extra work — particularly given the diversity of computing platforms and software versions (see also N. Barnes *Nature* **467,** 753; 2010).

These challenges also vary by discipline: analysis may comprise a few lines of code in some fields but thousands in others, as dictated by the requirements of individual papers.

Geoffrey J. Goodhill *University of Queensland, St Lucia, Australia.*
g.goodhill@uq.edu.au

Scottish separation could harm science

On 18 September, there will be a referendum in Scotland to decide whether it should become an independent country. As scientists and members of the campaign group Academics Better Together, we feel that Colin Macilwain gives an incomplete picture of Scottish science (*Nature* **493,** 579; 2013): we strongly believe that its brightest future is as part of the United Kingdom.

Scientists in Scotland benefit from being part of the large, efficient UK research community, in which competition and collaboration drive high-ranking research. As a small independent nation, Scotland would be forced to drop out of many research areas because it could no longer afford large-scale infrastructure. Collaborative research is likely to be more difficult across a national border. Also, Scotland would lose its disproportionately high block grant from the UK government, allocated in part to fund research and education in its universities.

The Scottish National Party's White Paper recognizes all this. The party wants Scotland to remain in the UK research-council system, but there are political indications that this may not be an option. The Wellcome Trust and the Association of Medical Research Charities have declared that it would be hard to fund research in another country. Scottish medical research would also be affected by restricted access to large UK clinical trials.

In our view, separating from the United Kingdom would leave scientists in Scotland with much to lose, for imperceptible gain.

Hugh Pennington *Aberdeen, UK.*
Susan A. Shaw *Dunblane, UK.*
Andrew Miller *Edinburgh, UK.*
miller_andrew1@sky.com

Former Iron Curtain safeguards wildlife

Aaron Ellison calls for conservation efforts across countries to offset the adverse effects of political borders on wildlife (*Nature* **508,** 9; 2014). The European Green Belt (www.europeangreenbelt.org), which is converting former cold-war territory in central and eastern Europe into a network of protected conservation areas, is one such initiative that should serve as a model for other regions with a history of strife.

The European Green Belt was instigated in 2003 and stretches 12,500 kilometres along the former Iron Curtain, the political barrier that existed from 1945 until 1989. The belt consists of protected core areas, sustainable-use areas, ecological corridors and buffer zones, which provide linked habitats and migration routes for such animals as wolves, bears and lynxes, as well as for amphibians and birds.

Not least, it offers a symbol of reconciliation, enjoying the patronage of Mikhail Gorbachev, former Soviet Union president.

Bernhard Jank, Johannes Rath *University of Vienna, Austria.*
bernhard.jank@outlook.com

Two brains and a forgotten theory

Two *Nature* papers published 100 years apart on the role of tension in brain cortex folding are connected by a historical footnote.

I discovered this as a result of a coincidence: the brain of the poet Walt Whitman and that of anatomist Andrew Parker ended up in a lab waste bin after dying within a week of one another in 1892. Both had been collected by the secretive American Anthropometric Society, which sought to uncover neuroanatomical features in the brains of eminent people. Unfortunately, Whitman's brain shattered after being dropped on the floor; Parker's crumbled after soaking in fixative for too long.

Beyond this shared fate of neural machinery, there is little information about Parker himself. My investigations revealed that he had proposed a mathematical theory of cortical folding based on the laws of liquid films and surface tension (see W. B. Benham *Nature* **55,** 619–620; 1897) — ironic, considering that the laws of liquid physics and chemical reactions crumbled his brain. I also found that his conclusions nicely complement a theory put forward a century later (D. C. Van Essen *Nature* **385,** 313–318; 1997).

Kevin S. Weiner *Stanford University; and the Institute for Applied Neuroscience, California, USA.*
kweiner@stanford.edu

CORRECTION

The data for the graphic in P. Ginsparg's Correspondence (*Nature* **508,** 44; 2014) were incorrectly credited to C. Labbé, who provided just a subset of the raw data used by the author.

ECOLOGY

Drought in the Congo Basin

A remote-sensing analysis of tropical forests in the Congo Basin that are experiencing chronic drought reveals consistent patterns of reduced vegetation greenness, increased temperatures and decreased water storage. SEE LETTER P.86

JEFFREY Q. CHAMBERS & DAR A. ROBERTS

The response of tropical forests to drought is a topic of considerable research and public-policy interest, and has been identified as a possible tipping point in Earth's climate system[1]. Drought in tropical forests has seasonal, episodic and chronic aspects; although seasonal and episodic events have been broadly studied for the Amazon tropical forests[2–8], chronic drought has received less attention, mainly because of the challenges involved in detecting long-term drought trends[9]. On page 86 of this issue, Zhou *et al.*[10] have significantly expanded the tropical-forest research programme by focusing on chronic drought in Africa's Congo Basin, a region that has been the subject of much less investigation than the Amazon.

Many tropical forests cannot be classified as classic 'rainforests' because they experience seasonal drought (generally defined as less than 100 millimetres of precipitation per month for 1–5 months per year). Trees in such forests are adapted to seasonal drought, and a few months of reduced precipitation have little effect on forest structure. By contrast, episodic events such as the Amazon mega-droughts of 2005 and 2010 can push tropical-forest trees outside their adaptive envelopes, resulting in mortality that is extensive enough to affect the atmospheric exchange of carbon, water and energy[4–6]. However, if the time between these episodic events is long enough, the forests may experience little overall change. Of most concern for tropical forests is chronic drought or frequent extreme drought, and climatic shifts to new, drier states.

Zhou and colleagues describe a study that uses optical, microwave and gravity remote-sensing data to evaluate long-term drought response in the Congo Basin (Fig. 1). Annual precipitation in this region is bimodal, and the authors focused on the second precipitation peak, which occurs between April and June. Rainfall data, including historical information from as far back as 1950 and more-recent data from the Tropical Rainfall Measuring Mission, reveal a steady decline in precipitation since 1985, and the authors report consistent observations of accompanying ecological changes. Reductions in the enhanced vegetation index (EVI; a measure of photosynthetic capacity, or 'greenness', obtained by the MODIS satellite-borne sensor) are reinforced by declines in vegetation optical depth (an indication of the leafy and woody components of living biomass), increases in land-surface temperature, declines in terrestrial water storage and changes to forest structure indicated by microwave backscatter. Increased photosynthetically active radiation (the amount of light available for photosynthesis) and decreased cloud optical thickness completed the picture of a tropical region experiencing decreased rainfall, decreased cloud cover and increased solar radiation for more than a decade.

MICHAEL FAY/NATIONAL GEOGRAPHIC CREATIVE

Figure 1 | Long-term drying. Tropical forests in the Congo Basin are experiencing chronic and increasing water shortage.

The consistency in all of these remotely sensed measures of drought contrasts with the more complex story in the Amazon. Field observations[5], land-surface temperatures[11] and radar-backscatter data[6] suggest declines in Amazon productivity during the 2005 mega-drought. And yet EVI analysis frequently indicated greening during Amazon mega-droughts[3,7], suggesting increased productivity[9]. These contrasting observations have led to considerable controversy regarding forest response to drought and potential artefacts in EVI.

Such artefacts stem from the fact that, although optical remote-sensing platforms measure changes in surface reflectance, the signal received by the sensor is also affected by atmospheric features, such as clouds and aerosols, and by changes in the orientation of the sensor with respect to the Sun (and associated shadowing). If the Sun is directly behind the sensor during imaging, shadows are minimized, and the shadow fraction increases at larger Sun–sensor angles. Such changes in Sun–sensor geometry can lead to false interpretations of changes in signal strength and drought response. For example, EVI has been shown to be highly sensitive to increased forest reflectivity, independent of changes in forest leaf area[2], and changes in EVI can be replicated by seasonal changes in Sun–sensor geometry[12].

Several factors may contribute to the differences between Zhou and colleagues' observations in the Congo and the more controversial reports from the Amazon.

One probable influence is the duration of the drought: Zhou *et al.* studied a chronic and increasingly severe water shortage, whereas whether the Amazon is also experiencing a long-term drying trend remains an open question[9]. However, analysis[8] using improved data for EVI and for the normalized difference vegetation index do in fact show large-scale 'browning' (reduction in greenness) in the Amazon in the mega-drought years of 2005 and 2010, and these observations are consistent with reduced microwave backscatter[6]. Thus, it seems plausible that the Amazon, like the Congo, has experienced large-scale structural responses to drought events, but that this was masked by remote-sensing artefacts.

Another crucial question is: what actually happens in the forest to cause these remotely sensed signals? The sensors generally respond to changes in the upper forest canopy, and those signals are not simple proxies for whole-ecosystem responses. To cause shifts in forest structure that drive climate-relevant atmospheric exchanges of carbon, water and energy, reductions in photosynthetic capacity must also cause other changes, such as reduced biomass production and elevated tree mortality.

One expected response to a long-term drying trend is a transition from high-biomass, closed-canopy forests to more-open, low-biomass forests and savannahs. However, the thresholds in water stress, carbon starvation, elevated temperature and increased vapour-pressure deficit at which this transition will occur are not well understood[13]. Response to drought is also not limited to upper-canopy effects, and other tools, such as tower-based measurements of evapotranspiration and net ecosystem productivity[14], coupled with field investigations of key ecosystem processes[15], are needed for complete assessments of the effects of drought on net forest–atmosphere fluxes.

Thus, a key constraint on our ability to interpret signals acquired by remote-sensing platforms is a lack of ground-based data with which to validate them. Obtaining such data will require extensive fieldwork using an array of methods at varying scales. As our climate continues to warm, quantifying the effects of drought on forests will become increasingly important, so ground-validated remote-sensing investigations must also be designed that best inform the development of Earth-system models. ■

Jeffrey Q. Chambers *is in the Department of Geography, University of California, Berkeley, Berkeley, California 94720-4740, USA.*
Dar A. Roberts *is in the Department of Geography, University of California, Santa Barbara, Santa Barbara, California 93106-4060, USA.*
e-mail: jqchambers@berkeley.edu

1. Lenton, T. M. *et al. Proc. Natl Acad. Sci. USA* **105,** 1786–1793 (2008).
2. Galvão, L. S. *et al. Remote Sens. Environ.* **115,** 2350–2359 (2011).
3. Huete, A. R. *et al. Geophys. Res. Lett.* **33,** L06405 (2006).
4. Marengo, J. A., Tomasella, J., Alves, L. M., Soares, W. R. & Rodriguez, D. A. *Geophys. Res. Lett.* **38,** L12703 (2011).
5. Phillips, O. L. *et al. Science* **323,** 1344–1347 (2009).
6. Saatchi, S. *et al. Proc. Natl Acad. Sci. USA* **110,** 565–570 (2013).
7. Samanta, A. *et al. Geophys. Res. Lett.* **37,** L05401 (2010).
8. Xu, L. *et al. Geophys. Res. Lett.* **38,** L07402 (2011).
9. Brando, P. M. *et al. Proc. Natl Acad. Sci. USA* **107,** 14685–14690 (2010).
10. Zhou, L. *et al. Nature* **509,** 86–90 (2014).
11. Toomey, M., Roberts, D. A., Still, C., Goulden, M. L. & McFadden, J. P. *Geophys. Res. Lett.* **38,** L19704 (2011).
12. Morton, D. C. *et al. Nature* **506,** 221–224 (2014).
13. McDowell, N. G. *et al. Trends Ecol. Evol.* **26,** 523–532 (2011).
14. Goulden, M. L. *et al. Ecol. Appl.* **14** (suppl.), 42–54 (2004).
15. Hutyra, L. R. *et al. J. Geophys. Res. Biogeosci.* **112,** G03008 (2007).

This article was published online on 23 April 2014.

APPLIED PHYSICS

Bright electron twisters

A new holographic method has been used to convert ordinary electron beams into helical beams. These beams show promise in applications such as the spectroscopic analysis of materials with intrinsic handedness and nanoparticle manipulation.

JUN YUAN

Electron vortex beams, also known as electron twisters, are quantum waves of electrons that have a rotational motion akin to that of the notional electron clouds orbiting atomic nuclei — except that there is no central Coulomb force to hold the electrons, so the quantum wave can travel along its axis of propagation as it rotates. These beams remained a theoretical novelty[1] until their experimental demonstration four years ago[2–4]. They have many interesting physical properties, including potentially high magnetic moments and their interaction with matter and electromagnetic fields[5]. To investigate them experimentally, the beams must be bright. In a paper published in *Applied Physics Letters,* Grillo *et al.*[6] show how this can be achieved.

In the framework of quantum mechanics, the rotational velocity of an electron vortex beam, and the related angular momentum, are proportional to the angular gradient of its quantum phase (a measure of the local amplitude of the quantum wave as it swings back and forth between its maximum and minimum values). Because of this phase gradient, the crest of the quantum wave rotates about the beam axis, tracing out a perfect helical wavefront. For comparison, lines of white froth on ocean waves approaching a beach mark the planar wavefronts of the waves.

The helical-wavefront nature of electron vortex beams suggests that they can be obtained by adding an angular phase gradient to an ordinary planar-wavefront electron beam. A technically challenging approach to making such a conversion is to use an optical element known as a phase plate whose thickness varies in a spiral fashion (a spiral-thickness profile)[2]. A more robust method[3,4] is to use a hologram, which was invented by physicist Dennis Gabor[7] in the late 1940s for reproducing three-dimensional images of objects. Holograms change the phase of an incoming light beam by altering the direction of travel of its wavefront by diffraction through a fringe-patterned mask.

First-generation holographic masks for producing electron vortex beams typically involve a simple 'black and white' (binary) fringe pattern, with the relative position of the fringes encoding information about the phase of the vortex beam. The 'white' areas of the mask are micro-machined away to allow the electron beam to pass through unimpeded, and the 'black' areas are filled with a beam-stopping material. However, the maximum efficiency of such a binary-amplitude hologram — defined as the percentage of the incident non-vortex beam that is converted into a vortex beam of a particular orbital angular momentum — is just some 10% (ref. 6). This is because not only is about half of the non-vortex beam stopped by the black areas, but also the remaining transmitted half is distributed among many diffracted beams.

Enter Grillo and colleagues. The authors report a device for creating electron vortex beams with an efficiency of 25%. They attained this value by combining the advantages of the holographic and phase-plate approaches. In this hybrid approach, the conventional binary mask is replaced by one that does not contain beam-stopping areas and which has a sawtooth-thickness structure (Fig. 1) that acts as a phase plate and directs the electrons preferentially in one direction. In optical science, such a device is called a blazed phase hologram, and has a theoretical efficiency approaching 100%.

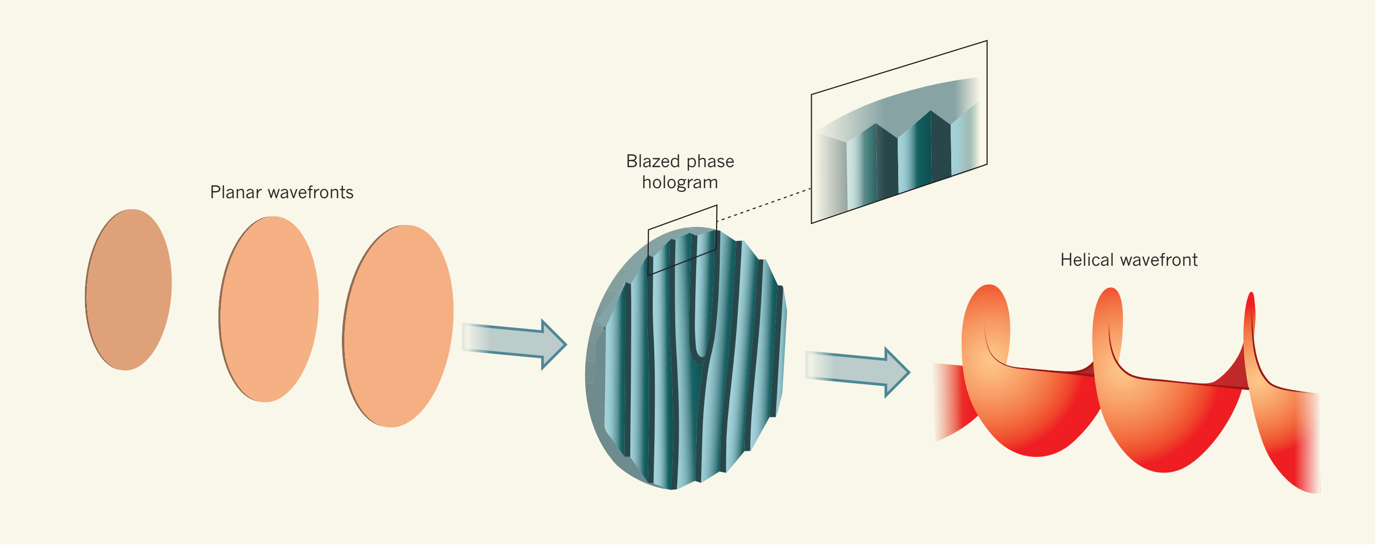


Figure 1 | Converting wavefronts. Grillo *et al.*[6] have designed a blazed phase hologram that has a sawtooth-thickness structure. The device converts an electron beam in which the wavefronts form planes into a beam with a wave crest that rotates about its axis of propagation, tracing out a helical wavefront.

But there are caveats to this approach. Obtaining bright electron vortex beams by this method is challenging because it requires the thickness profile of the blazed phase hologram to be controlled with nanometre-level accuracy. Also, part of the electron beam passing through the hologram will inevitably lose energy through a process called inelastic scattering, which leads to a non-vortex background signal. For beam diagnostics, this inelastic component of the beam can be removed using energy-filtering methods. However, the use of purely phase-shifting devices, such as those that exploit an effect known as optical aberration[8], instead of a blazed phase hologram, might be preferable for applications such as spectroscopy based on the chirality (handedness) of the vortex beams.

The helical form of an electron vortex beam's wavefront means that the exact phase of the beam is ill-defined at its centre, resulting in a doughnut-shaped beam-intensity structure that can be less than 1 nanometre in diameter[9]. This length scale is about 1,000 times smaller than that of existing optical vortex beams, which are used to trap and move micrometre-sized particles. Bright electron vortex beams produced using Grillo and colleagues' method may therefore allow nanoparticles and even individual atoms to be easily manipulated. In fact, existing, rather 'dim' electron vortex beams have already been used to transfer orbital angular momentum from the beams to nanoparticles[3,10,11].

The authors' method will also allow the production of bright electron vortex beams of very high orbital angular momentum, which will enable the investigation of subtle quantum effects associated with the giant magnetic moments of such beams. Finally, owing to the beams' intrinsic chirality, intense electron vortex beams could be used for the spectroscopic study of chiral materials[3,12], such as magnetic materials, certain polymers and biological macromolecules. The future of electron vortex beams is undoubtedly getting brighter. ■

Jun Yuan *is in the Department of Physics, University of York, York YO10 5DD, UK.*
e-mail: jun.yuan@york.ac.uk

1. Bliokh, K. Yu., Bliokh, Y. P., Savel'ev, S. & Nori, F. *Phys. Rev. Lett.* **99,** 190404 (2007).
2. Uchida, M. & Tonomura, A. *Nature* **464,** 737–739 (2010).
3. Verbeeck, J., Tian, H. & Schattschneider, P. *Nature* **467,** 301–304 (2010).
4. McMorran, B. J. *et al. Science* **331,** 192–195 (2011).
5. Lloyd, S. M., Babiker, M., Yuan, J. & Kerr-Edwards, C. *Phys. Rev. Lett.* **109,** 254801 (2012).
6. Grillo, V. *et al. Appl. Phys. Lett.* **104,** 043109 (2014).
7. Gabor, D. *Nature* **161,** 777–778 (1948).
8. Clark, L. *et al. Phys. Rev. Lett.* **111,** 064801 (2013).
9. Idrobo, J. C. & Pennycook, S. J. *J. Electron Microsc.* **60,** 295–300 (2011).
10. Verbeeck, J., Tian, H. & Van Tendeloo, G. *Adv. Mater.* **25,** 1114–1117 (2013).
11. Gnanavel, T., Yuan, J. & Babiker, M. *Proc. Eur. Microsc. Congr.* www.emc2012.org.uk//documents/Abstracts/Abstracts/EMC2012_1082.pdf (2012).
12. Yuan, J., Lloyd, S. M. & Babiker, M. *Phys. Rev. A* **88,** 031801 (2013).

NEUROSCIENCE

Feedback throttled down for smooth moves

A group of regulatory neurons in the spinal cord has been found to reduce sensory feedback to muscles in mice. Removal of these neurons leads to repetitive limb oscillations during reaching. **SEE ARTICLE P.43**

STEPHEN H. SCOTT
& FRÉDÉRIC CREVECOEUR

Sensory signals from our limbs allow us to interpret a wealth of information, from perceiving the objects we touch to correcting errors during movement. But despite their importance, the signals are turned down (throttled down) when we move[1]. How does this happen, and why? On page 43 of this issue, Fink *et al.*[2] report that, in mice, the signals are throttled down by a set of neurons in the spinal cord, and that removal of these neurons causes the animals' limbs to oscillate dramatically whenever they reach for food.

Although motor control involves many pathways and circuits in the spinal cord and brain, Fink and colleagues' study focused on the simplest: the feedback between muscle sensory afferent neurons (which carry impulses from the muscle towards the spinal cord) and efferent motor neurons (which carry signals from the spinal cord to the muscles; Fig. 1a). Your doctor examines this pathway when she or he taps your tendon: contact between the hammer and the tendon excites sensory afferents in the stretched muscle, and the impulses are then transferred from the axon terminal

at the end of the afferent to the motor neuron, across the sensory–motor synapse — the junction between the two neurons. The feedback leads to muscle activity and a flinching movement at the joint.

Most neural feedback pathways involve intermediary neurons, called interneurons, which form connections with the motor neurons after the synapse, and permit substantial processing of signals. But the direct projection from sensory afferents to motor neurons precludes such processing. Instead, the activity of these synapses (and other afferent synapses in the spinal cord) is regulated before the synapse. In these cases, the axon terminal of the sensory afferent is contacted by the terminal of another axon, from a subgroup of GABAergic interneurons — named for the γ-aminobutyric acid (GABA) neurotransmitter they release. The 'axo-axonic' synapse at this junction can inhibit the afferent axon terminal, diminishing the amount of neurotransmitter released across the sensory–motor synapse (Fig. 1b). This reduces the feedback gain: the ratio between sensory-afferent input and motor-neuron output.

Presynaptic inhibition of sensory signals has been shown to correlate with voluntary motor actions, such as reaching for objects[1,3]. The mechanism is thought to filter incoming sensory information and help the brain to extract relevant information[1,4]. However, why sensory feedback to control movement should also be reduced at this time remains unclear.

Unravelling this mystery required several steps. First, Fink *et al.* used a strain of mice designed to enable labelling of a subgroup of GABAergic interneurons in the spinal cord to make the interneurons express the proteins Channelrhodopsin-2 and yellow fluorescent protein. Examination of how yellow fluorescent protein was distributed in the spinal cord showed that these neurons almost exclusively create axo-axonic synapses on afferent axon terminals. Next, the researchers took advantage of the fact that Channelrhodopsin-2 causes the interneurons to fire when exposed to light. Light activation reduced transmission across the sensory–motor synapse, demonstrating that this specific group of GABAergic neurons generates presynaptic inhibition.

Finally, Fink and colleagues removed the GABAergic interneurons from the motor circuits of the mice, and analysed motor function in the mutants. This required another genetic trick — the authors forced the neurons to express a toxin-receptor protein, which killed the neurons when the toxin was injected into the spinal cord. The mice could still walk after this treatment, even across a ladder, and they could also maintain stationary body postures. However, whenever they reached towards food, their forelimbs oscillated dramatically. This suggests that presynaptic inhibition has a key role in maintaining smooth voluntary movements.

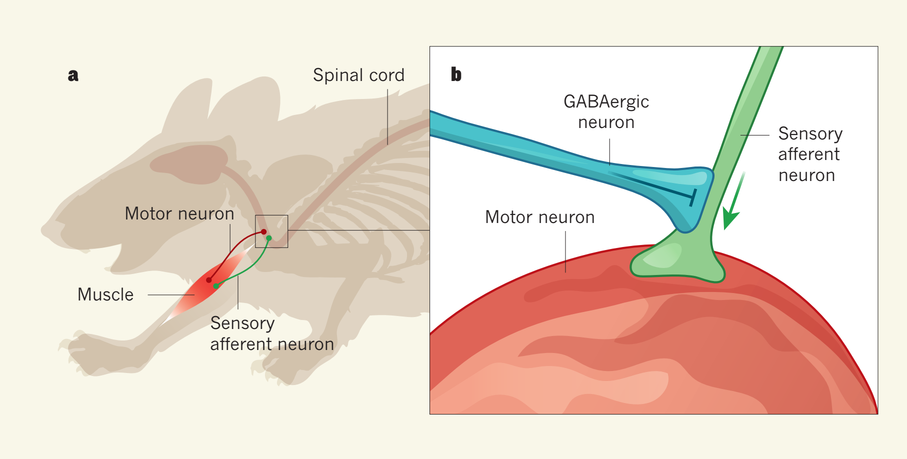


Figure 1 | Controlling reaching movements in mice. **a**, When muscle is stretched, sensory afferent neurons are stimulated, and impulses travel towards the spinal cord. A sensory–motor feedback pathway causes motor neurons to transmit impulses back to the muscles, generating muscle activity and movement at the joint. **b**, A close-up of the sensory–motor synapse, where the end of the afferent neuron (the axon terminal) contacts the motor neuron. Fink *et al.*[2] report that a subtype of regulatory GABAergic interneuron in the spinal cord contacts the afferent axon terminal and, when active, diminishes the amount of neurotransmitter molecules released across the sensory–motor synapse, thereby inhibiting feedback to motor neurons.

Why does the forelimb oscillate when presynaptic inhibition is removed? Oscillations occurred during reaching, but not when animals were maintaining a fixed posture or walking, suggesting that there is something special about voluntary movement that suddenly makes feedback to motor neurons too high. Fink and co-workers suggest a simple interpretation, based on a pair of opposing muscles pulling at a joint: when feedback is too strong they become reciprocally active, oscillating the limb.

There are, however, other interpretations of this result that deserve attention. For instance, when muscles are forced to lengthen, they initially generate a large resistive force (high stiffness), which quickly recedes as lengthening proceeds, leaving a smaller resistive force (low stiffness)[5,6]. The same is true for muscle shortening during movement[7]. These muscle properties influence feedback control, so the feedback gains that provide smooth corrections when animals are maintaining a fixed posture could potentially generate oscillations during voluntary movement. Alternatively, oscillations may arise during voluntary movement because of an imbalance between spinal and brain feedback. In any case, presynaptic inhibition seems to throttle down spinal feedback to ensure smooth movement — a clever biological trick!

The need to reduce sensory feedback during movement seems to be at odds with a growing body of evidence[8] showing its powerful contribution to the control of voluntary movement. This apparent paradox is reconciled by evidence[9,10] suggesting that the brain supports more-complex sensorimotor processing than the spinal cord, so high-gain control is processed through brain pathways. Faster, direct spinal feedback may be used in some behaviours, but in others it must be reduced through presynaptic inhibition.

The past few years have seen an explosion of molecular and genetic techniques that have allowed us to identify different types of neuron, artificially modify their activity and remove them from neural circuits. For example, removal of the V2a spinal neuron, which projects to motor neurons and indirectly to the cerebellum, impairs reaching, but does not generate large oscillations[11]. Together, these techniques provide a powerful set of tools to tease apart spinal (and brain) circuits, leading to a more-complete understanding of how these distributed networks support smooth and highly flexible motor actions. ■

Stephen H. Scott *and* **Frédéric Crevecoeur** *are in the Department of Biomedical and Molecular Sciences, Centre for Neuroscience Studies, Queen's University, Kingston, Ontario K7L 3N6, Canada.*
e-mail: steve.scott@queensu.ca

1. Rudomin, P. *Exp. Brain Res.* **196,** 139–151 (2009).
2. Fink, A. J. P. *et al. Nature* **509,** 43–48 (2014).
3. Seki, K., Perlmutter, S. I. & Fetz, E. E. *Nature Neurosci.* **6,** 1309–1316 (2003).
4. Wolpert, D. M., Diedrichsen, J. & Flanagan, J. R. *Nature Rev. Neurosci.* **12,** 739–751 (2011).
5. Rack, P. M. H. & Westbury, D. R. *J. Physiol.* **240,** 331–350 (1974).
6. van Eesbeek, S., de Groot, J. H., van der Helm, F. C. T. & de Vlugt, E. *J. Biomech.* **43,** 2539–2547 (2010).
7. Axelson, H. W. & Hagbarth, K. E. *J. Physiol. (Lond.)* **535,** 279–288 (2001).
8. Scott, S. H. *Trends Cogn. Sci.* **16,** 541–549 (2012).
9. Pruszynski, J. A. *et al. Nature* **478,** 387–390 (2011).
10. Franklin, D. W. & Wolpert, D. M. *Neuron* **72,** 425–442 (2011).
11. Azim, E., Jiang, J., Alstermark, B. & Jessell, T. M. *Nature* http://dx.doi.org/10.1038/nature13021 (2014).

GEOPHYSICS

Making the Earth move

Controversy exists over the roles of water and melt in the ductile layer of the mantle beneath Earth's surface plates. New data support models in which small amounts of melting occur in the uppermost part of this region. SEE LETTER P.81

ROB L. EVANS

In plate tectonics, the process by which Earth's surface is constantly being reorganized and rebuilt, surface plates (the lithosphere) move over the underlying part of the mantle that is actively transporting heat by convection (the asthenosphere). Almost since plate tectonics was discovered, there has been debate over the nature of the uppermost asthenosphere, particularly whether a lubricating mechanism is required to weaken it and facilitate plate motion. Numerical-modelling studies that attempt to reproduce the geometry and behaviour of plates suggest that some mechanism that lowers the viscosity of a boundary layer beneath the plates is indeed necessary[1]. Into this debate comes the paper by Sifré *et al.*[2] on page 81 of this issue, in which the authors argue that a small amount of melting of the asthenosphere is consistent with geophysical observations and the presence of active volcanoes on old sea floor (Fig. 1).

Unravelling the mechanisms responsible for weakening the asthenosphere is relevant not only to finding out how tectonics operates on Earth, but also to understanding the conditions under which it might operate on other planets. Two leading hypotheses invoked to explain how the viscosity of the asthenosphere might be reduced are a damp composition[3] (primarily the result of hydrogen dissolved in the mineral olivine) and some degree of melting. To a certain extent, these two mechanisms work together: a high water content promotes melting, but water is preferentially removed and apportioned into the melt — if there is too much melting, this layer of the mantle dries out, raising the viscosity and also potentially shutting off further melting. Sifré and colleagues focus on the melting hypothesis in the laboratory, by measuring the electrical conductivity of melts typically produced in the asthenosphere.

Earth's electrical conductivity is sensitive both to the presence of connected networks of melt and to water content. This latter effect has been the source of much controversy in the past decade or so, with laboratory measurements showing considerable disagreement over the magnitude of possible enhanced conductivity due to water in olivine[4] and hence over whether water can explain conductivity models derived from field observations. Models that favour a small amount of melting over a high water content have also come under fire, because the amount of melt needed to explain both conductivity and seismic-velocity field observations has been thought to be too high for the melt to remain *in situ*. Also questioned is whether melt can act as a suitable weakening mechanism[5].

Sifré and colleagues' conductivity measurements suggest that previous estimates of mantle conductivity in areas in which melting is just starting have probably greatly overestimated melt fractions. This result comes from studying the conductivity of silicate melts that formed at the onset of melting in a mantle containing water and carbon dioxide, both of which promote melting at lower temperatures than would be the case in their absence[6]. It has been known for some years that the conductivity of silicate melts depends on their composition[7,8], and more recent results have highlighted how sodium can increase silicate-melt conductivity[9]. But the authors' data show a significant enhancement in conductivity for hydrous silicate melts rich in CO_2. These data allow models of asthenospheric conductivity to be interpreted in terms of very small amounts of melt, possibly sufficient to lubricate the upper asthenosphere and facilitate plate motion while remaining in the asthenosphere. In this model, the base of the lithosphere becomes a freezing boundary, with melt above it being unstable. Importantly, melt conductivity is shown to be a function of the amount of melting.

This is a step forward, but it complicates the interpretation of field observations: geophysicists will have to pay careful attention when applying petrological models to their data, particularly in terms of how water and CO_2 are distributed between minerals before and during melting. They will also have to figure out how to untangle the effects of melt and water when analysing field data. Sifré *et al.* lay out a range of models for the oceanic lithosphere and asthenosphere that are eminently testable through observations. Some observations have already been made, and no doubt a reanalysis of some existing data is in order. Whether the authors' models will stand the test of time remains to be seen. For example, it might be argued that the CO_2 contents considered by the authors are unrealistically high.

Recent seismic observations[10] paint a picture of the mantle for the whole Pacific basin that also helps to clarify how the oceanic lithosphere evolved. Although these observations do not resolve the primary argument of whether the asthenosphere is wet or partially molten (or both), they do define structural boundaries within which models such as those presented by Sifré and colleagues must be assessed. These seismic data seem to suggest that, at sea-floor ages greater than about

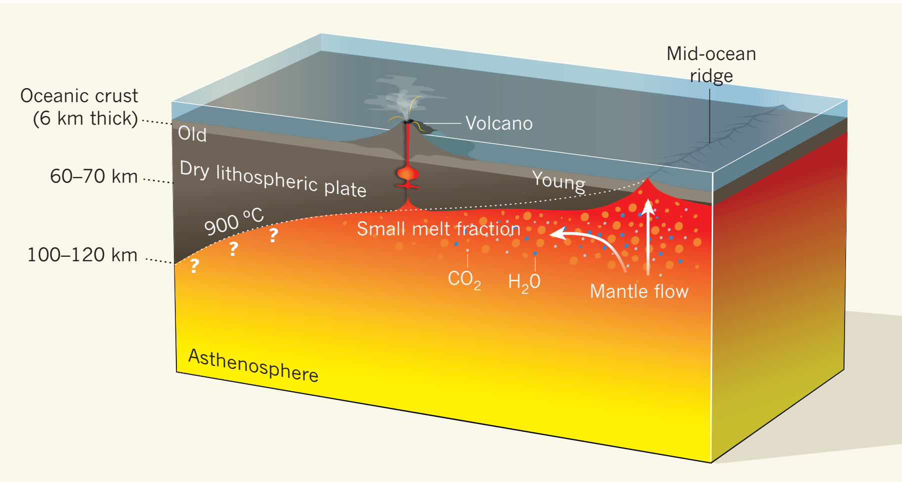


Figure 1 | A conceptual model of Earth's oceanic lithosphere and asthenosphere. In this model[2,3,9,12–14], melting processes resulting from mantle upwelling beneath a mid-ocean ridge produce a dry lithospheric plate. This plate rides above an asthenospheric mantle layer that is damp and contains some carbon dioxide, generating a small fraction of mantle melt (some of which escapes through sea-floor volcanoes), as suggested by Sifré and colleagues[2]. The combination of water and melt weakens the layer, facilitating plate motion (not shown). At greater lithospheric ages, thermal structure becomes more significant and a section of the upper asthenosphere cools, stiffens and accretes onto the base of the lithosphere; the dotted line roughly denotes the 900 °C isotherm. The evolution of H_2O and CO_2 content, the presence of melt and the vertical transition from lithosphere to asthenosphere are not well constrained by existing field observations, nor is the subsequent evolution of old lithosphere (question marks).

70 million years, the thickness and thermal structure of the lithosphere are greater than those for the compositional and melting boundaries considered by Sifré and co-workers. But even if the specifics of the authors' models turn out to be incorrect, particularly the depths of various melting boundaries, the core of their assertion — that melt conductivities are much more complex than previously imagined — must be taken into account.

These results suggest that field measurements of asthenospheric conductivity have the potential to be significantly more sensitive to the presence of incipient melts than seismological studies, because conductivity is influenced by markedly lower melt fractions than is seismic velocity. Thus, although low melt fractions can explain observed conductivities, the same may not be true of low seismic velocities. In fact, a water-driven mechanism has been invoked[11] to describe the sharp decrease in velocities observed across the base of the lithosphere.

Having geophysical techniques of differing sensitivities is important, and probably offers a means of unravelling the competing effects of water and melt if multiple methods are used together. Indeed, several efforts that take a multi-technique approach are already under way, although a complete model for all seafloor ages is a long way off. But such a model is crucial if we are to understand the evolution of lithospheric and asthenospheric structure. The large pool of instruments available worldwide means that our ability to make such observations has never been better. With the new crop of laboratory data in hand, including those of Sifré and co-workers, we should be well positioned to understand what allows Earth's surface plates to move so freely. ■

Rob L. Evans *is in the Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.*
e-mail: revans@whoi.edu

1. Tackley, P. J. *Geochem. Geophys. Geosyst.* **1,** 2000GC000043 (2000).
2. Sifré, D. *et al. Nature* **509,** 81–85 (2014).
3. Hirth, G. & Kohlstedt, D. L. *Earth Planet. Sci. Lett.* **144,** 93–108 (1996).
4. Evans, R. L. in *The Magnetotelluric Method: Theory and Practice* (eds Chave, A. D. & Jones, A. G.) Ch. 3A, 50–95 (Cambridge Univ. Press, 2012).
5. Karato, S. *Earth Planet. Sci. Lett.* **321–322,** 95–103 (2012).
6. Dasgupta, R., Hirschmann, M. M. & Smith, N. D. *Geology* **35,** 135–138 (2007).
7. Ni, H., Keppler, H. & Behrens, H. *Contrib. Mineral. Petrol.* **162,** 637–650 (2011).
8. Pommier, A. & Le-Trong, E. *Comput. Geosci.* **37,** 1450–1459 (2011).
9. Pommier, A. & Garnero, E. J. in *AGU Fall Meet.*, San Francisco, Abstr. MR24B-02 (2013).
10. Beghein, C., Yuan, K., Schmerr, N. & Xing, Z. *Science* **343,** 1237–1240 (2014).
11. Olugboji, T. M., Karato, S. & Park, J. *Geochem. Geophys. Geosyst.* **14,** 880–901 (2013).
12. Evans, R. L. *et al. Nature* **437,** 249–252 (2005).
13. Gaherty, J. B., Jordan, T. H. & Lind, S. G. *J. Geophys. Res.* **101,** 22291–22309 (1996).
14. Naif, S., Key, K., Constable, S. & Evans, R. L. *Nature* **495,** 356–359 (2013).

ASTRONOMY

A new spin on exoplanets

Spectroscopic observations of a young exoplanet have allowed its spin velocity to be directly measured. Its fast spin is in accord with the extrapolation of the known trend in spin velocity as a function of planet mass. SEE LETTER P.63

TRAVIS BARMAN

Planetary spin is something we notice daily on Earth, as the Sun, Moon, planets and stars rise and set. Of the planets in our Solar System, five of them spin rapidly and in the same direction as they orbit the Sun. This is unlikely to have occurred by chance, and reflects important, but poorly understood, aspects of how the planets formed. Although the discovery of planets orbiting stars other than the Sun has expanded our view of planetary sizes and masses, little is known about the diversity of planetary spins. On page 63 of this issue, Snellen *et al.*[1] report how they have, for the first time, directly measured the spin of an exoplanet called β Pictoris b (ref. 2). The result paves the way for measuring the spins of a large sample of exoplanets.

Most exoplanets are hidden behind the much brighter light from the star they orbit, making detailed study difficult. Recently, however, several gas-giant planets have been directly imaged, with planet and star light clearly separated — β Pictoris b is one example (Fig. 1). Direct imaging is easiest for young gas-giant planets, which are still hot and bright from their recent formation. β Pictoris b is very young[3] (less than 1% of the age of the Sun), more than 10,000 times brighter than Jupiter, only about 20 parsecs (65 light years) from Earth, and has a separation from its parent

50 Years Ago

If many zoologists and plant pathologists have given scant attention to the soil-inhabiting and plant-parasitic nematodes, the main reason must be that they are difficult to handle. This is due partly to an awkward size and shape and partly to a marked fragility before mechanical and osmotic forces, so that implements and reagents must be used with care and skill. Hence the appearance in 1949, as the Ministry of Agriculture and Fisheries *Technical Bulletin* No. 2, of *Laboratory Methods for Work with Plant and Soil Nematodes* … New editions followed in 1951 and … 1957, and recently a fourth edition has been published … This justly reflects the increasing interest in the group now taken by workers in many countries.
From *Nature* 2 May 1964

100 Years Ago

The April number of *Science Progress* contains an editorial article of nine pages, entitled "Sweating the Scientist." During the past year an inquiry has been conducted by our contemporary as to the emoluments of scientific workers … As might, perhaps, have been anticipated, the replies received suffice to prove the "low scale of payment given throughout the British Empire for such work." … The article is a timely one, and deserves the careful attention of all scientific workers, as the question of remuneration is one of paramount importance to the future welfare of science in this country. Particular reference is made to the unpaid services of men of science upon Government Committees, and to the custom of Government departments going to learned societies for expert advice for which no payment is made. "In other words, the State exploits the man of science on account of his enthusiasm for his work and his patriotism."
From *Nature* 30 April 1914

70 million years, the thickness and thermal structure of the lithosphere are greater than those for the compositional and melting boundaries considered by Sifré and co-workers. But even if the specifics of the authors' models turn out to be incorrect, particularly the depths of various melting boundaries, the core of their assertion — that melt conductivities are much more complex than previously imagined — must be taken into account.

These results suggest that field measurements of asthenospheric conductivity have the potential to be significantly more sensitive to the presence of incipient melts than seismological studies, because conductivity is influenced by markedly lower melt fractions than is seismic velocity. Thus, although low melt fractions can explain observed conductivities, the same may not be true of low seismic velocities. In fact, a water-driven mechanism has been invoked[11] to describe the sharp decrease in velocities observed across the base of the lithosphere.

Having geophysical techniques of differing sensitivities is important, and probably offers a means of unravelling the competing effects of water and melt if multiple methods are used together. Indeed, several efforts that take a multi-technique approach are already under way, although a complete model for all seafloor ages is a long way off. But such a model is crucial if we are to understand the evolution of lithospheric and asthenospheric structure. The large pool of instruments available worldwide means that our ability to make such observations has never been better. With the new crop of laboratory data in hand, including those of Sifré and co-workers, we should be well positioned to understand what allows Earth's surface plates to move so freely. ■

Rob L. Evans *is in the Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.*
e-mail: revans@whoi.edu

1. Tackley, P. J. *Geochem. Geophys. Geosyst.* **1,** 2000GC000043 (2000).
2. Sifré, D. *et al. Nature* **509,** 81–85 (2014).
3. Hirth, G. & Kohlstedt, D. L. *Earth Planet. Sci. Lett.* **144,** 93–108 (1996).
4. Evans, R. L. in *The Magnetotelluric Method: Theory and Practice* (eds Chave, A. D. & Jones, A. G.) Ch. 3A, 50–95 (Cambridge Univ. Press, 2012).
5. Karato, S. *Earth Planet. Sci. Lett.* **321–322,** 95–103 (2012).
6. Dasgupta, R., Hirschmann, M. M. & Smith, N. D. *Geology* **35,** 135–138 (2007).
7. Ni, H., Keppler, H. & Behrens, H. *Contrib. Mineral. Petrol.* **162,** 637–650 (2011).
8. Pommier, A. & Le-Trong, E. *Comput. Geosci.* **37,** 1450–1459 (2011).
9. Pommier, A. & Garnero, E. J. in *AGU Fall Meet.*, San Francisco, Abstr. MR24B-02 (2013).
10. Beghein, C., Yuan, K., Schmerr, N. & Xing, Z. *Science* **343,** 1237–1240 (2014).
11. Olugboji, T. M., Karato, S. & Park, J. *Geochem. Geophys. Geosyst.* **14,** 880–901 (2013).
12. Evans, R. L. *et al. Nature* **437,** 249–252 (2005).
13. Gaherty, J. B., Jordan, T. H. & Lind, S. G. *J. Geophys. Res.* **101,** 22291–22309 (1996).
14. Naif, S., Key, K., Constable, S. & Evans, R. L. *Nature* **495,** 356–359 (2013).

ASTRONOMY

A new spin on exoplanets

Spectroscopic observations of a young exoplanet have allowed its spin velocity to be directly measured. Its fast spin is in accord with the extrapolation of the known trend in spin velocity as a function of planet mass. SEE LETTER P.63

TRAVIS BARMAN

Planetary spin is something we notice daily on Earth, as the Sun, Moon, planets and stars rise and set. Of the planets in our Solar System, five of them spin rapidly and in the same direction as they orbit the Sun. This is unlikely to have occurred by chance, and reflects important, but poorly understood, aspects of how the planets formed. Although the discovery of planets orbiting stars other than the Sun has expanded our view of planetary sizes and masses, little is known about the diversity of planetary spins. On page 63 of this issue, Snellen *et al.*[1] report how they have, for the first time, directly measured the spin of an exoplanet called β Pictoris b (ref. 2). The result paves the way for measuring the spins of a large sample of exoplanets.

Most exoplanets are hidden behind the much brighter light from the star they orbit, making detailed study difficult. Recently, however, several gas-giant planets have been directly imaged, with planet and star light clearly separated — β Pictoris b is one example (Fig. 1). Direct imaging is easiest for young gas-giant planets, which are still hot and bright from their recent formation. β Pictoris b is very young[3] (less than 1% of the age of the Sun), more than 10,000 times brighter than Jupiter, only about 20 parsecs (65 light years) from Earth, and has a separation from its parent

NATURE

50 Years Ago

If many zoologists and plant pathologists have given scant attention to the soil-inhabiting and plant-parasitic nematodes, the main reason must be that they are difficult to handle. This is due partly to an awkward size and shape and partly to a marked fragility before mechanical and osmotic forces, so that implements and reagents must be used with care and skill. Hence the appearance in 1949, as the Ministry of Agriculture and Fisheries *Technical Bulletin* No. 2, of *Laboratory Methods for Work with Plant and Soil Nematodes* … New editions followed in 1951 and … 1957, and recently a fourth edition has been published … This justly reflects the increasing interest in the group now taken by workers in many countries.
From *Nature* 2 May 1964

100 Years Ago

The April number of *Science Progress* contains an editorial article of nine pages, entitled "Sweating the Scientist." During the past year an inquiry has been conducted by our contemporary as to the emoluments of scientific workers … As might, perhaps, have been anticipated, the replies received suffice to prove the "low scale of payment given throughout the British Empire for such work." … The article is a timely one, and deserves the careful attention of all scientific workers, as the question of remuneration is one of paramount importance to the future welfare of science in this country. Particular reference is made to the unpaid services of men of science upon Government Committees, and to the custom of Government departments going to learned societies for expert advice for which no payment is made. "In other words, the State exploits the man of science on account of his enthusiasm for his work and his patriotism."
From *Nature* 30 April 1914

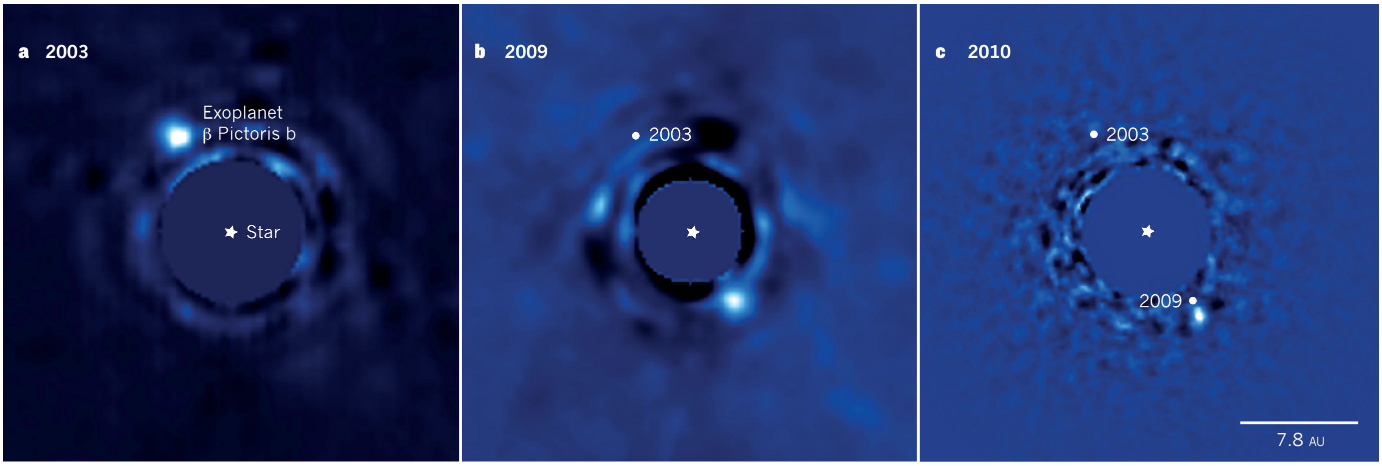


Figure 1 | Movement of exoplanet β Pictoris b around its host star. Infrared images of the planet β Pictoris b obtained in 2003 (**a**), 2009 (**b**) and 2010 (**c**), showing the planet's movement in an orbital plane that is nearly edge-on as seen from Earth. The host star is in the central part, but its light has been suppressed to show the fainter planet. The white dots in **b** and **c** denote previous positions of the planet. Faint blobs are optical effects. It is not possible to tell from these images whether the planet is orbiting towards or away from us, but Snellen and colleagues' spectroscopic observations[1] clearly indicate that the planet is currently in a part of its orbit where it is moving towards us. (1 astronomical unit (AU) is the mean Earth–Sun distance.)

star roughly twice the Jupiter–Sun distance These characteristics make it an ideal target for modern spectrographs.

Spectroscopy is often used to determine the exact velocities of stars and the variations induced by orbiting planets. In Snellen and colleagues' study, it is the orbital and rotational velocities of the planet, not of its host star, that are directly measured with this method. The technique leverages the distinct spectral features produced by molecules found in the atmospheres of giant planets but not in host stars as hot as that of β Pictoris b. The molecules present depend on the elemental composition, temperature and pressure of the planet's atmosphere, but, in general, water, carbon monoxide, ammonia and methane are the most common. Each of these produces a rich spectrum of absorption features that is different from the fairly smooth spectrum of the star, which contains only a few atomic absorption lines.

These planetary molecular 'fingerprints' are shifted in wavelength relative to the star's spectrum by an amount that depends on the orbital velocity and orbital inclination of the planet. In addition to this wavelength shift, the planet's spin broadens the planetary spectral lines, because half of the planet rotates towards the observer, shifting the spectrum towards bluer wavelengths, and the other half rotates away, redshifting the spectrum. The combined red- and blueshifts cause spectral lines to become shallower and broader depending on the speed of rotation and the inclination of the spin axis with respect to Earth.

Snellen and collaborators clearly detected a shifted carbon monoxide spectral signature at a wavelength corresponding to an orbital velocity consistent with estimates of β Pictoris b's orbital period[4]. The carbon monoxide signature is blueshifted, telling us that the planet is currently orbiting towards Earth — information not obtainable by imaging observations alone (Fig. 1). The authors also found that this carbon monoxide signal is broadened by a rotational velocity greater than, or equal to, 25 kilometres per second — which is about 50 times faster than that of Earth and twice as fast as Jupiter's. β Pictoris b has a much larger diameter than Earth and its spin results in a rotation period (the length of its day) of about 8 hours.

In cold gas-giant planets, methane is by far the most abundant carbon-bearing molecule. However, β Pictoris b is more than 10 times hotter than Jupiter, and carbon chemistry favours carbon monoxide at high temperatures[5], so its presence in the planet's spectrum was no surprise. It has also been recognized for some time that the carbon monoxide lines are useful for velocity measurements of low-mass objects, including warm planets, because they stand out cleanly from neighbouring spectral features caused by water absorption.

The fast spin of this planet is also not out of the ordinary. Generally speaking, the Solar System planets show a trend, with spin increasing with increasing mass (see Fig. 2 of the paper[1]). The trend does not hold for Mercury and Venus; both have undergone significant spinning down caused by gravitational interactions with the Sun and other perturbations. Interestingly, at its young age, β Pictoris b is still contracting as it cools and will eventually shrink to a radius comparable to that of Jupiter, leading to an even faster spin when it reaches the age of the Sun. As pointed out by Snellen and colleagues, this eventual spin velocity agrees remarkably well with the Solar System trend.

The mass–spin relation in the Solar System is intriguing given the different compositions and formation scenarios of the rocky and gas-giant planets. It may be related to the fact that many of our planets spin within factors of about 2 to 10 of their break-up velocity (as does β Pictoris b), and this velocity is a function of mass and radius. The trend may also be evidence that impacts with large bodies, which impart significant angular momentum to the planet, may have a crucial role in the late stages of planet formation across the full range of planet masses. Some caution is warranted, however, because β Pictoris b's large mass (roughly 10 times that of Jupiter) gives it much in common with brown dwarfs, objects that have masses between those of stars and planets and spins[6] that are equally consistent with that of β Pictoris b. So, any similarities with Solar System trends may simply be a coincidence.

Nevertheless, Snellen and colleagues' determination of a young exoplanet's spin will shed light on the early conditions of planet formation. Such spin measurements are also relevant for understanding the dominant meteorological processes on exoplanets such as β Pictoris b, where a fast spin implies that its atmospheric dynamics at large length scales is mainly influenced by its rotation[7]. The number of directly imaged giant planets will soon increase substantially, and many of these planets will be excellent targets for spin measurements in the near future. ■

Travis Barman *is at the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721-0092, USA.*
e-mail: barman@lpl.arizona.edu

1. Snellen, I. A. G. *et al. Nature* **509,** 63–65 (2014).
2. Lagrange, A.-M. *et al. Science* **329,** 57–59 (2010).
3. Binks, A. & Jeffries, R. D. *Mon. Not. R. Astron. Soc.* **438,** L11–L15 (2014).
4. Macintosh, B. *et al.* Preprint at http://arxiv.org/abs/1403.7520 (2014).
5. Lodders, K. & Fegley, B. *Icarus* **155,** 393–424 (2002).
6. Reiners, A. & Basri, G. *Astrophys. J.* **684,** 1390–1403 (2008).
7. Showman, A. P. & Kaspi, Y. *Astrophys. J.* **776,** 85–104 (2013).
8. Bonnefoy, M. *et al. Astron. Astrophys.* **528,** L15 (2011).

ARTICLE

doi:10.1038/nature13276

Presynaptic inhibition of spinal sensory feedback ensures smooth movement

Andrew J. P. Fink[1], Katherine R. Croce[1], Z. Josh Huang[2], L. F. Abbott[3], Thomas M. Jessell[1] & Eiman Azim[1]

The precision of skilled movement depends on sensory feedback and its refinement by local inhibitory microcircuits. One specialized set of spinal GABAergic interneurons forms axo–axonic contacts with the central terminals of sensory afferents, exerting presynaptic inhibitory control over sensory–motor transmission. The inability to achieve selective access to the GABAergic neurons responsible for this unorthodox inhibitory mechanism has left unresolved the contribution of presynaptic inhibition to motor behaviour. We used *Gad2* as a genetic entry point to manipulate the interneurons that contact sensory terminals, and show that activation of these interneurons in mice elicits the defining physiological characteristics of presynaptic inhibition. Selective genetic ablation of *Gad2*-expressing interneurons severely perturbs goal-directed reaching movements, uncovering a pronounced and stereotypic forelimb motor oscillation, the core features of which are captured by modelling the consequences of sensory feedback at high gain. Our findings define the neural substrate of a genetically hardwired gain control system crucial for the smooth execution of movement.

Animals execute skilled limb movements with seemingly effortless precision, belying an elaborate series of neural transformations that direct each motor act. Spinal motor output relies on local inhibitory interneurons that shape the response of motor neurons to diverse excitatory inputs[1]. Most inhibitory interneurons form direct postsynaptic connections with motor or premotor neurons[2], but a small subset of GABAergic interneurons instead forms axo–axonic contacts with sensory afferent terminals, regulating sensory–motor drive through a process of presynaptic inhibition (Fig. 1a)[3,4]. This presynaptic inhibitory system can be recruited by sensory, descending and local neuronal inputs[4–6], implying its pivotal role in the control of motor output. However, despite the occurrence of axo–axonic contacts at most sensory terminals[7,8], the predominance of postsynaptic inhibition has left unresolved the motor behavioural significance of this presynaptic control system.

Presynaptic inhibition has been characterized most extensively at proprioceptive sensory–motor synapses[3,4,6]. Proprioceptors convey the state of muscle contraction to motor neurons, through direct and indirect feedback pathways[1,9]. Elimination of proprioceptive feedback impairs

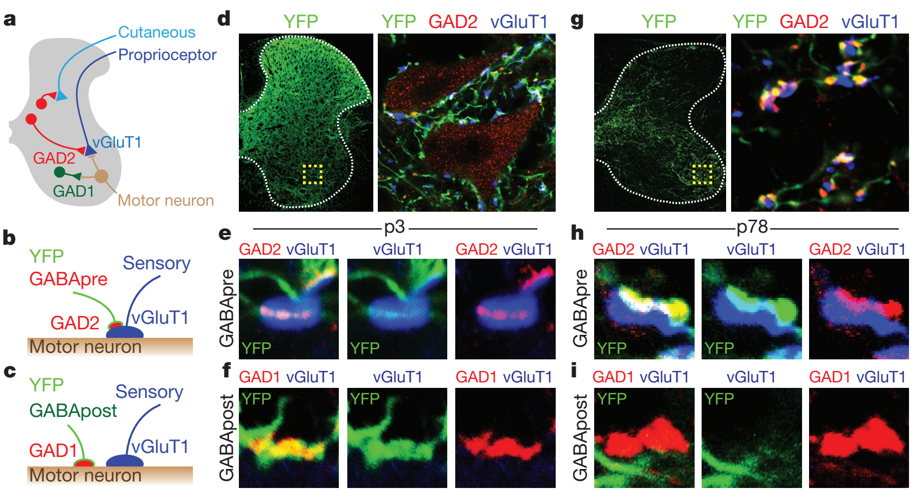


Figure 1 | Genetic targeting of GABApre neurons. **a**, GABApre contacts on sensory terminals; GABApost contacts on motor neurons. **b**, GABApre, $GAD2^+/GAD1^+$ boutons on $vGluT1^+$ proprioceptor terminals. **c**, GABApost, $GAD2^{OFF}/GAD1^+$ boutons not on $vGluT1^+$ terminals. **d–f**, Postnatal day (p) 3 *FLEX-ChR2-YFP* lumbar injection in $Gad2^{Cre}$ mice. **d**, ChR2–YFP expression in lumbar spinal cord. In ventral horn (box, shown on right) $YFP^+/GAD2^+$ GABApre boutons contact $vGluT1^+$ proprioceptor terminals (**e**, higher magnification); **f**, $YFP^+/GAD1^+$ GABApost boutons do not contact $vGluT1^+$ terminals. p0–3 injection marks GABApre (85.2 ± 2.1%) and GABApost (57.3 ± 3.4%; $n = 3$) boutons. **g–i**, p78 *FLEX-ChR2-YFP* cervical injection in $Gad2^{Cre}$ mice. **g**, ChR2–YFP expression in cervical spinal cord. In ventral horn (box, shown on right) $YFP^+/GAD2^+$ GABApre boutons contact $vGluT1^+$ terminals (**h**, higher magnification); **i**, $YFP^{OFF}/GAD1^+$ GABApost bouton. Adult injection marks GABApre (77.6 ± 8.5%) but not GABApost (1.0 ± 0.1%; $n = 2$) boutons. Values, mean ± s.e.m.

[1]Howard Hughes Medical Institute, Kavli Institute for Brain Science, Mortimer B. Zuckerman Mind Brain Behavior Institute, Departments of Neuroscience and Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [3]Center for Theoretical Neuroscience, Departments of Physiology and Neuroscience, Columbia University, New York, New York 10032, USA.

motor coordination[10], establishing a basal requirement for sensory transmission in motor control. Conversely, limiting the gain of proprioceptive feedback has been proposed, on theoretical grounds, to be a critical determinant of motor stability[11]. In principle, the divisive nature of presynaptic inhibition provides an effective means of controlling sensory gain[9,12–14], but without a way to manipulate the relevant set of inhibitory interneurons it has not been possible to resolve whether, or how, presynaptic inhibition contributes to motor behaviour.

The inhibitory interneurons that form axo–axonic contacts with sensory terminals differ from other spinal GABAergic neurons in that they alone express GAD2 (also known as GAD65), one of two GABA-synthetic enzymes (Fig. 1a–c)[7,15,16]. We have used *Gad2* as a genetic entry point for manipulating presynaptic inhibitory interneurons in mice and assessing their role in motor behaviour. Our findings indicate that *Gad2*-expressing interneurons mediate presynaptic inhibition at sensory–motor synapses, and that selective elimination of these interneurons elicits motor oscillations during goal-directed reaching. The essential features of this motor perturbation can be captured by a simple model in which high-gain proprioceptive feedback induces limb oscillation. This alignment of behaviour and theory argues that presynaptic inhibitory interneurons ensure the smooth operation of goal-directed motor behaviours by modulating the gain of sensory feedback.

Targeting GABApre neurons

We first asked whether the GABAergic interneurons that contact sensory terminals (GABApre neurons) can be defined by transgene reporter expression in a *Gad2-IRES-Cre* (*Gad2*^{Cre}) mouse line[17]. To mark *Gad2*-expressing interneurons (*Gad2*^{Cre} neurons) we injected a recombinant adeno-associated viral vector encoding a Cre-recombinase-dependent (*FLEX*) channelrhodopsin (ChR2)–yellow fluorescent protein (YFP) fusion protein into the spinal cord of *Gad2*^{Cre} mice.

For physiological studies of presynaptic inhibition we targeted *Gad2*^{Cre} neurons in early postnatal lumbar spinal cord (Supplementary Note 1)[18]. At this stage *Gad2* is expressed both by GABApre neurons as well as by GABApost neurons that contact motor neurons and premotor interneurons (Fig. 1a–c)[7,19]. Injection of *FLEX-ChR2-YFP* at postnatal day (p) 0–3 into lumbar segments resulted, 14–21 days later, in broad YFP expression with dense axonal labelling in the vicinity of motor neurons (Fig. 1d). At this stage 36% of GABAergic terminals near motor neurons derive from GABApre neurons, and 64% from GABApost neurons. We found that 85% of GABApre and 57% of GABApost boutons in the ventral spinal cord expressed ChR2–YFP (Fig. 1b–f). Thus early postnatal *FLEX-ChR2-YFP* injection marks GABApre and GABApost boutons at similar incidence.

For motor behavioural studies we targeted *Gad2*^{Cre} neurons in adult cervical spinal cord (Supplementary Note 1 and Extended Data Fig. 1). Cervical injection of *FLEX-ChR2-YFP* at ~p78, with analysis 14–21 days later, revealed that 78% of GABApre and ~1% of GABApost boutons expressed YFP (Fig. 1b, c, g–i). Thus adult *Gad2*^{Cre} transduction marks GABApre neurons in a near-selective manner.

Gad2^{Cre}-mediated presynaptic inhibition

We determined whether activation of *Gad2*^{Cre} interneurons elicits the two hallmarks of presynaptic inhibition: primary afferent depolarization and suppression of sensory neurotransmitter release[4].

We first examined whether photoactivation of ChR2-expressing *Gad2*^{Cre} neurons entrains neuronal spiking. Recordings from ChR2–YFP^{+} neurons in p9−14 lumbar spinal cord preparations from *Gad2*^{Cre} mice injected with *FLEX-ChR2-YFP* at p0–3 (Fig. 2a) revealed that photostimulation (473 nm, ~10 mW) elicited action potentials that followed frequencies up to ~50 Hz (Fig. 2b–d)[20]. Thus targeted ChR2 expression confers optical control of *Gad2*^{Cre} neuronal spiking.

We then determined whether GABA released upon *Gad2*^{Cre} neuronal photoactivation depolarizes sensory neurons, eliciting primary afferent depolarization (PAD; Fig. 2e)[21]. Single-pulse photoactivation of *Gad2*^{Cre} interneurons elicited PAD with an amplitude and time course similar to that induced by dorsal root simulation (Fig. 2e, f). Antagonists of GABA-A receptors, but not glycine receptors, abolished *Gad2*^{Cre}

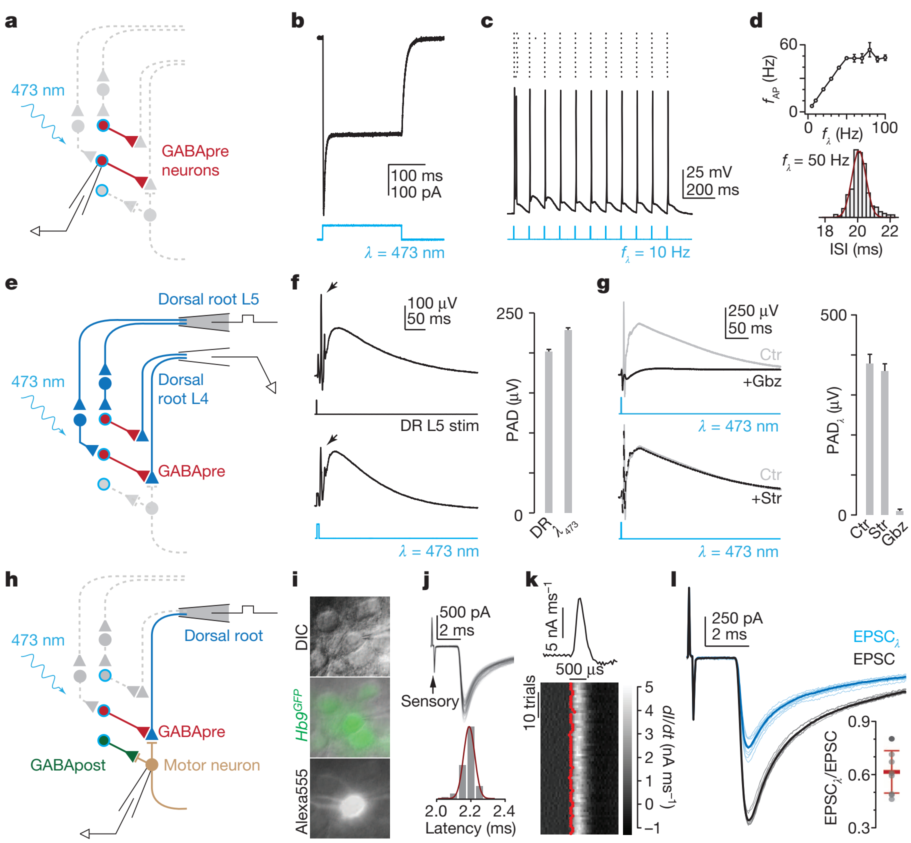


Figure 2 | *Gad2*^{Cre} photoactivation elicits presynaptic inhibition. **a**, Recording from ChR2–YFP^{+} (blue outline) *Gad2*^{Cre} neurons (red). **b–d**, Photostimulation (λ) induced currents (**b**) and action potentials (**c**), whose frequency (f_{AP}, **d**) followed light-pulse frequency (f_{λ}) up to ~50 Hz (top, $n = 3$). Distribution of interspike intervals (ISI) for $f_{\lambda} = 50$ Hz (bottom, 19.9 ± 0.6 ms s.d., $n = 500$ trials). **e**, Dorsal root stimulation (L5) and recording (L4). **f**, L5-evoked (top, DR) and light-evoked (bottom, λ_{473}; $n = 4$) PAD. Arrows, antidromic spikes. **g**, Light-evoked PAD without (Ctr) and with SR 95531 (Gbz, 2 µM) or strychnine (Str, 5 µM). Plots (**f**, **g**) show PAD amplitude. **h**, Isolating sensory input during *Gad2*^{Cre} photoactivation. **i**, Motor neuron identified using differential interference contrast (DIC) optics, *Hb9*^{GFP} expression and Alexa^{555} fill. **j**, **k**, Monosynaptic sensory-EPSCs (**j**; 90 trials, 10 Hz, arrow indicates sensory stimulation); onset latencies (Gaussian fit, red) estimated using EPSC waveform derivative (**k**; top, mean; bottom, individual traces; EPSC onsets, red). Mean onset latency 2.87 ± 0.18 ms; cv_{onset} 0.018 ± 0.002; $n = 19$. **l**, Sensory-EPSCs (0.1 Hz; mean, bold; raw, faint) without (EPSC, black) or with photoactivation ($EPSC_{\lambda}$, blue; fifteen 1 ms light pulses at 50 Hz; 45 ms delay). Inset, EPSC suppression; $n = 9$. Values and error bars, mean ± s.e.m., except **l**, mean ± s.d. See Extended Data Figs 2, 4 for additional quantification.

neuron-evoked PAD (Fig. 2g), establishing its GABAergic character[22]. However, PAD reflects the depolarization of cutaneous as well as proprioceptive afferents[21], prompting us to ask whether *Gad2*^{Cre} neuron activation depolarizes proprioceptive afferents. At reduced temperatures *in vivo* PAD evokes transmitter release from proprioceptor terminals, depolarizing motor neurons (Supplementary Note 2)[21]. Consistent with this, we found that *Gad2*^{Cre} photoactivation *in vitro* at 24–26 °C elicited a GABA-A and AMPA receptor-dependent motor neuron depolarization (Extended Data Fig. 2). Activation of *Gad2*^{Cre} neurons therefore elicits PAD at proprioceptor terminals.

To assess the impact of *Gad2*^{Cre} neuronal activation on sensory–motor transmission we isolated sensory input to motor neurons. Stimulation of individual L3 to L5 dorsal roots elicited monosynaptic excitatory postsynaptic currents (EPSCs) in motor neurons (mean amplitude 1.2 ± 0.3 nA s.e.m.; coefficient of variation of onset (cv_{onset}) < 0.02; $n = 19$; Fig. 2h–k and Extended Data Fig. 3)[23]. Photoactivation of *Gad2*^{Cre} neurons elicited a frequency-dependent reduction in sensory-evoked EPSC amplitude that persisted for >800 ms (Fig. 2l, inset: mean ~40%, range 19–53% reduction in EPSC amplitude; two-tailed paired *t* test, $P < 10^{-4}$, $n = 9$; Extended Data Fig. 4). Thus synchronous activation of *Gad2*^{Cre} neurons elicits a long-lasting suppression of sensory-evoked EPSCs (Supplementary Note 3).

At the neonatal stages used for physiological analysis *Gad2*^{Cre} marks both GABApre and GABApost neurons. Nevertheless, three findings indicate that *Gad2*^{Cre} suppression of sensory-evoked EPSCs reflects presynaptic inhibition. First, *Gad2*^{Cre}-evoked inhibitory postsynaptic currents (*Gad2*^{Cre}-IPSCs) persisted for only ~30 ms (Fig. 3a–c and Extended Data Fig. 5a–c), whereas *Gad2*^{Cre} EPSC suppression lasted long after *Gad2*^{Cre}-IPSCs had decayed (Extended Data Figs 4a and 5a, b)[4]. Second, *Gad2*^{Cre} EPSC suppression is exclusively GABAergic, whereas coexpression of glycine by most GABApost neurons underlies the fact that *Gad2*^{Cre}-IPSCs are predominantly glycinergic (Fig. 3a–g)[24,25]. Third, *Gad2*^{Cre} EPSC suppression did not alter EPSC waveforms, arguing against the idea that the reduction in EPSC amplitude reflects an increased motor neuron membrane conductance (Extended Data Fig. 5d–j). These findings reveal that GABApost neurons make only a minimal contribution to the inhibition of sensory-evoked EPSCs, thus implicating GABApre neurons.

The second core feature of presynaptic inhibition is a reduced probability of sensory transmitter release[26,27]. Modelling shows that this reduction can be captured by the kinetics of decay in amplitude of successive sensory-evoked EPSCs (Fig. 3h, i)[28,29]. We found that upon repetitive dorsal root stimulation, photoactivation of *Gad2*^{Cre} neurons elicited a reduction in the rate of EPSC amplitude decay that was readily fit by a decrease in the release probability parameter of a short-term depression model ($p_r = 0.69 \pm 0.03$, control; 0.50 ± 0.05 s.e.m., photoactivation; two-tailed paired *t* test, $P < 0.001$, $n = 9$; Fig. 3i, j and Extended Data Fig. 6)[29]. The fidelity of this fit indicates that activation of *Gad2*^{Cre} neurons reduces sensory transmitter release probability. Therefore *Gad2*^{Cre} neurons mediate both classical features of presynaptic inhibition at sensory–motor synapses.

Limb oscillation after GABApre ablation

In primates presynaptic inhibition has been implicated in the control of forelimb movement[30], prompting us to use goal-directed reaching in mice as a behavioural routine for examining the impact of removing presynaptic inhibition (Supplementary Note 1). We eliminated GABApre neurons by injecting an adeno-associated virus encoding a Cre-recombinase-dependent diphtheria toxin receptor-green fluorescent protein fusion (*FLEX-DTR-GFP*) unilaterally into C3–T1 spinal segments in p56–84 *Gad2*^{Cre} mice (Fig. 4a)[31]. Administration of diphtheria toxin (DT) 14–21 days later resulted in a >90% depletion of GAD2$^+$/GAD1$^+$ boutons on ventral vGluT1$^+$ sensory terminals, without altering the number of GABApost GAD2^{OFF}/GAD1$^+$ boutons (Fig. 4b). Thus adult *Gad2*^{Cre}-based DTR targeting achieves selective elimination of GABApre neurons. Mice lacking GABApre neurons exhibited two prominent behavioural phenotypes: impaired reaching and forelimb scratching.

To analyse forelimb motor behaviour, mice were trained to reach for and retrieve a food pellet located on the far side of a narrow access window[31,32], with reaches analysed before and after GABApre neuronal ablation (Fig. 4c, d). DT administration elicited a marked degradation in reach accuracy, such that by the seventh post-DT day fewer than 5% of trials were successful, compared to a pre-DT success rate of ~50% (Fig. 4e). We detected no change in left-right forelimb alternation, nor in the accuracy of forepaw placement in a horizontal ladder stepping task (Fig. 4f)[31,33]. Although these findings do not preclude a role for presynaptic inhibition in locomotor control, they do indicate a degree of task-selectivity in the degradation of motor performance.

To evaluate forelimb movement we quantified limb kinematics by tracking a reflective marker attached to the right forepaw (Fig. 4d)[31], comparing reach trajectory and velocity in individual mice before and after GABApre ablation. Before DT administration mice displayed smooth paw trajectories that varied little across trials or with trial success (Extended Data Fig. 7a and Supplementary Videos 1 and 2)[31]. In contrast, after GABApre ablation the stereotypically smooth nature of reaches was lost and paw trajectories exhibited frequent reversals in direction (Fig. 5a–c,

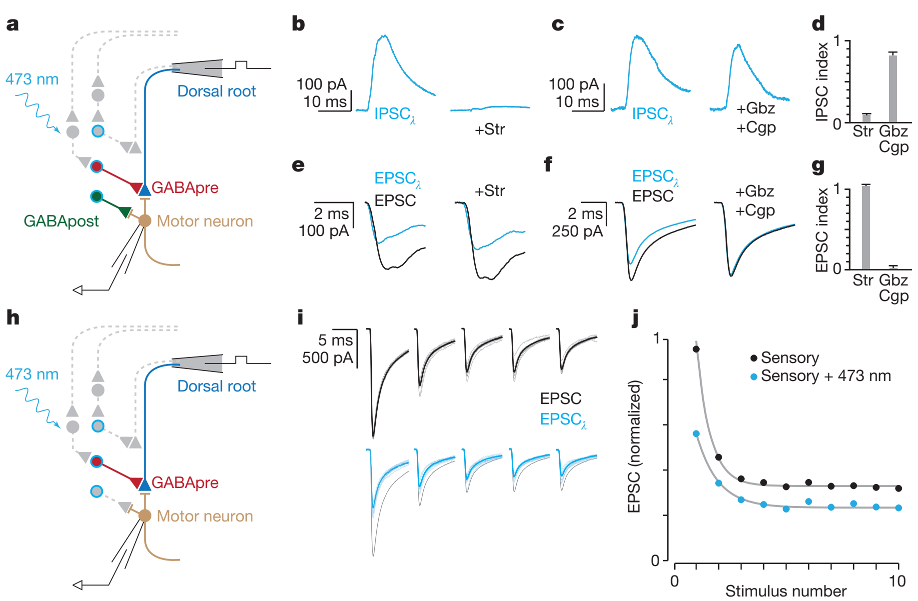


Figure 3 | GABApre photoactivation reduces sensory transmitter release. **a**, Distinguishing GABApre from GABApost inhibition. **b–g**, Pharmacological analysis. **b–d**, Light-evoked (λ) IPSCs plus strychnine (Str; $n = 5$) (**b**), SR 95531 (Gbz; $n = 3$) and CGP 54626 (Cgp, 2.5 μM; $n = 3$) (**c**); IPSC index, ratio of IPSC amplitude with drug to control (**d**). **e–g**, *Gad2*^{Cre} suppression of sensory-EPSCs (with photostimulation, blue; without, black) with strychnine (Str, $n = 3$) (**e**); SR 95531 and CGP 54626 (Gbz/Cgp, $n = 4$) (**f**); EPSC index, no effect = 1; drug abolishes EPSC-suppression = 0 (see Methods for description and Extended Data Fig. 5 for statistics) (**g**). **h**, Isolation of GABApre inhibitory action. **i**, Sensory-EPSCs (10 pulses, 25 Hz; mean, bold; raw, faint) without (black) or with photostimulation (blue, with black control superimposed for comparison). **j**, Normalized mean EPSC amplitudes from **i**. Grey, exponential fits to model-generated EPSC amplitudes. Values and error bars, mean ± s.e.m.

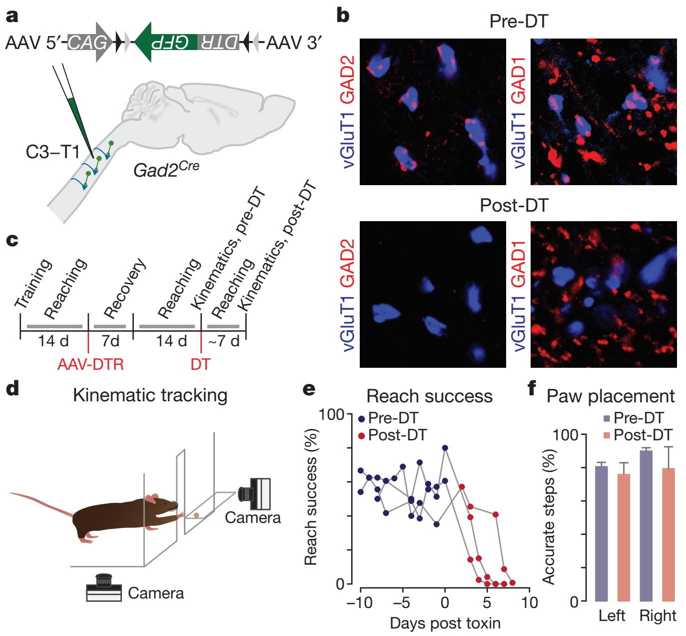


Figure 4 | Selective GABApre ablation. **a**, After C3−T1 *FLEX-DTR-GFP* injection into adult *Gad2*^{Cre} mice, **b**, GABApre boutons (GAD2^{+}, red) contact proprioceptor terminals (vGluT1^{+}, blue) pre- but not post-DT administration. Only GAD1-marked GABApost boutons remain post-DT. vGluT1^{+} boutons with GABApre contacts, pre-DT 88 ± 2% (n = 2,058 boutons in 4 mice); post-DT 9 ± 1% (n = 4,347 boutons in 3 mice). **c**, Experimental timeline. **d**, Reaching kinematics assay (see Methods)[31]. **e**, Reach success pre-DT and post-DT (3 mice shown). **f**, Stepping accuracy, horizontal ladder assay (n = 4)[33]. Values and error bars, mean ± s.e.m. See Extended Data Fig. 9 and Extended Data Table 1 for statistics.

Extended Data Fig. 7b, Extended Data Table 1 and Supplementary Videos 3 and 4). Kinematic defects were restricted to the 'reach' phase — before the paw passes through the access window (Fig. 5c and Extended Data Table 1). We observed no perturbation of the limb at rest; during the initiation of forelimb movement; during the late 'grab' phase — when the paw pronates in anticipation of pellet grasp; nor in digit abduction during grasp attempts (Fig. 5b–e, Extended Data Table 1 and Supplementary Video 5; see Methods and Supplementary Discussion).

To assess the periodicity of aberrant forelimb movements upon GABApre ablation we examined the temporal structure of forepaw kinematics. The spectral power of pre-DT reaches decayed steadily with increasing oscillation frequency (Fig. 5f, g). In contrast, post-DT reach oscillations exhibited an ~7.4-fold increase in peak spectral amplitude within the 18–22 Hz frequency range, focused at 19.5 ± 0.5 Hz s.e.m., with little variability across reaches and animals ($cv_{frequency}$ = 0.06, n = 5; Fig. 5f, g). We also detected a sub-harmonic series with spectral peaks at 10.8 ± 0.5 Hz and 5.9 ± 0.3 Hz (Fig. 5f). The decay of post-DT oscillations could be fit by a single exponential function with a time constant of 77 ± 3 ms s.e.m., again consistent across animals (cv_{decay} = 0.08, n = 5; Fig. 5h, i). Thus elimination of GABApre neurons uncovers a structured oscillation of constant frequency and decay time, defining features of a damped harmonic oscillator.

A GABApre gain control system

Prior models of sensory–motor control have proposed that increasing the gain of proprioceptive feedback elicits motor oscillation[11]. Can a simple model of sensory–motor feedback gain provide a theoretical framework for interpreting the oscillations that follow GABApre ablation? We simulated a simplified forelimb joint controlled by flexor and extensor torques and subject to gain-modulated sensory feedback, fitting the frequency and decay time of the model's single oscillatory eigenmode to experimentally-derived post-DT values (Fig. 6a, b, Extended Data Fig. 8a, b and Supplementary Note 4). At low feedback gain this model displayed smooth joint extension, resembling pre-DT reach velocity profiles (Fig. 6a, c; compare with Fig. 5d, top). At high gain, simulating the loss of divisive (presynaptic) inhibition, joint extension was dominated by a cycle of flexor-extensor alternation that resembled post-DT reach velocity profiles (Fig. 6b, d; compare with Fig. 5d, bottom). Reducing subtractive (postsynaptic) inhibition did not result in similar oscillations (Fig. 6c, green line, Supplementary Note 4 and Supplementary Discussion). This theoretical analysis suggests that oscillations produced by elimination of presynaptic inhibition cannot be prevented by remaining intact postsynaptic inhibition, consistent with our experimental observations.

How do joint oscillations evolve as gain increases? Across a wide range of gain levels, the oscillatory frequency and decay times were constrained to a narrow set of values (Fig. 6e, f and Supplementary Note 4), providing a potential basis for the constancy of motor oscillation observed across mice. Moreover, similar oscillations emerged across multiple feedback delays (Extended Data Fig. 8c–h), accommodating the potential involvement of direct and indirect sensory feedback circuits (Supplementary Note 5). Thus a simple model of proprioceptive feedback-driven oscillation can account for the principal motor deficits that emerge upon elimination of presynaptic inhibition.

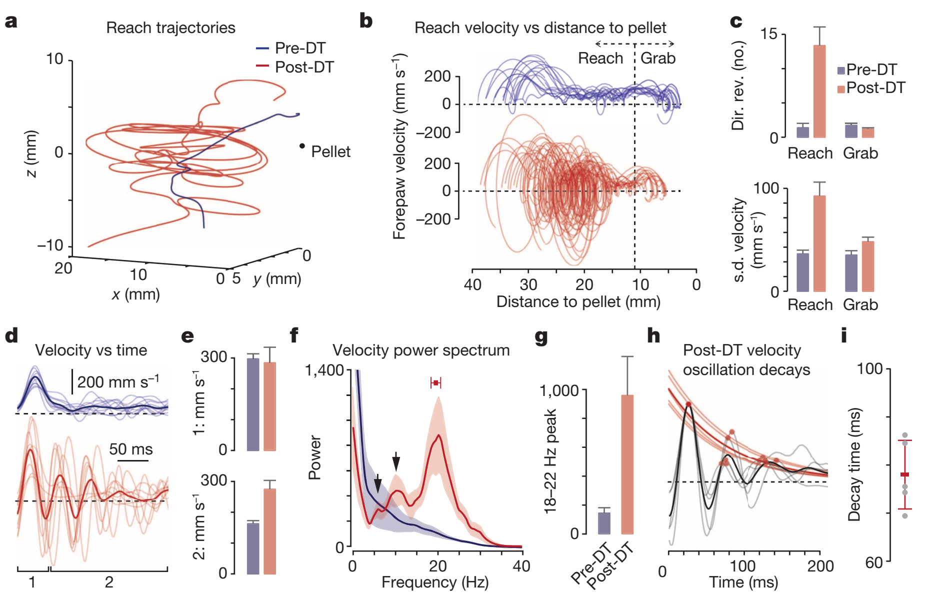


Figure 5 | GABApre ablation uncovers limb oscillation during reaching. **a**, **b**, Reach trajectories (**a**) and velocities (**b**) from a representative mouse. Vertical line, reach versus grab phases. Horizontal line, forepaw direction reversals (also in **d**, **h**). **c**, Number of direction reversals (Dir. rev.) and s.d. of velocity between reaches during reach and grab phases (n = 4 mice, 112 reaches). **d**, Paw velocity versus time from a second representative mouse. Mean, bold; individual reaches, faint. **e**, Peak reach velocities: during initiation (bracket '1' in **d**) and subsequent movement (>50 ms, bracket '2' in **d**) (n = 4). **f**, Mean power spectra of reach velocities; shaded area, s.e.m. Red box, mean frequency peak ± s.d. Arrows, subharmonic peaks. **g**, Peak power, 18–22 Hz (n = 5 mice, 132 reaches; see Methods). **h**, Peak-aligned averages within (grey) and across mice (black). Mean velocity peaks for each mouse (red dots) and exponential fits (faint red lines). Mean fit, bold red line. **i**, Decay time constant across mice (mean ± s.d.; n = 5).Values and error bars, mean ± s.e.m., except where indicated. See Extended Data Table 1 for statistics.

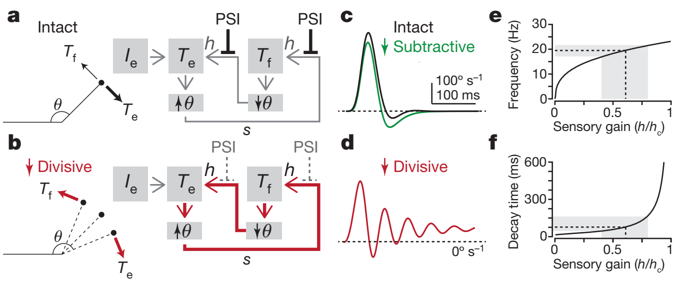


Figure 6 | GABApre neurons regulate sensory gain. **a**, **b**, Model of joint angle θ under the control of flexor T_f and extensor T_e torques. External drive I_e activates T_e, increasing θ (bold arrow). Sensory feedback s recruits a counteracting T_f with efficacy specified by sensory gain h. Presynaptic inhibition (PSI) suppresses gain divisively, keeping T_f small (faint arrow, **a**). With reduced PSI, high-gain feedback induces alternating, sensory-driven T_f and T_e (red arrows, **b**). **c**, **d**, Joint angle velocity over time with intact ($h/h_c = 0.1$, low gain) and reduced subtractive (postsynaptic) inhibition (green line, see Methods) (**c**), and with reduced divisive PSI ($h/h_c = 0.6$, high gain) (**d**). For gains above the critical level (h_c), oscillations do not decay. **e**, **f**, Oscillation frequency (**e**) and decay time (**f**) versus gain. Grey boxes, range of gain values ($h/h_c = 0.4$–0.8) corresponding to post-DT oscillation frequencies and decay times (Fig. 5d–i). Dashed lines, 19.5 Hz, 77 ms. See Supplementary Note 4.

Finally, we considered whether this propriocentric view of feedback-driven oscillation is challenged by the forelimb scratching behaviour observed after GABApre ablation (Extended Data Fig. 9a, b). In particular, we asked whether defects in forelimb reach could reflect enhanced cutaneous input. We applied the Na^+ channel blocker lidocaine topically to the forelimb of GABApre-ablated mice, with the intent of selectively inactivating cutaneous sensory endings. Lidocaine eliminated scratching behaviour but left reaching defects and the ~20 Hz forelimb oscillation unaffected (Extended Data Fig. 9c–f and Supplementary Video 6). These results argue against the notion that enhanced cutaneous gain contributes to motor oscillation and aberrant reaching. Our findings therefore place GABApre neurons at the core of a proprioceptive gain control system for suppression of oscillations during goal-directed forelimb movement.

Discussion

Despite long appreciation of presynaptic inhibitory control, its contribution to motor behaviour has remained unclear. We have identified, activated and eliminated the spinal GABAergic interneurons responsible for presynaptic inhibition at sensory synapses, revealing their central role in suppressing motor oscillation during forelimb movement. Our findings address three core issues in sensory–motor control: the behavioural relevance of sensory gain scaling; the neural underpinnings of smooth limb trajectory; and the grain of neural circuitry that engages presynaptic inhibition.

The rich repertoire of mammalian limb movements demands that proprioceptive afferents sustain a wide range of firing frequencies as they strive to supply task-appropriate feedback[34]. The dynamic character of sensory firing implies in principle that the strength of presynaptic inhibition should scale with the intensity of sensory input[35,36]. Several strategies underlie the dynamic scaling of sensory feedback gain by GABApre neurons. At a circuit level GABApre neurons can be driven to high firing rates by proprioceptive input[37], and our physiological studies reveal a strict proportionality between the frequency of GABApre neuronal spiking and the strength of presynaptic inhibition. These two features suggest that GABApre neurons scale sensory input gain in register with proprioceptor firing frequency. At a synaptic level the prolonged time-course of presynaptic inhibition has the consequence that gain changes endure despite rapid fluctuation in sensory firing frequency over the course of a movement (Supplementary Note 3). The rapid spiking sustained by GABApre neurons also provides insight into the selectivity of expression of GAD2: by virtue of its association with synaptic vesicles GAD2 facilitates GABA release at high firing frequencies[16,38]. Thus both the microcircuitry and neurochemistry of GABApre neurons contribute to the adaptive scaling of sensory feedback during motor control.

Motor circuits throughout the central nervous system are prone to oscillation[39–41]. Yet the smoothness that normally characterizes limb trajectories implies that oscillations are suppressed before motor neuron activation[42]. Behavioural studies, in conjunction with modelling, suggest that the GABApre gain control system provides one effective means of suppressing forelimb oscillation. We observe that mice at rest lack an overt tremor, consistent with the idea that oscillations are driven by proprioceptive feedback, a central tenet of the gain model. Moreover, the shallow relationship between gain and oscillation revealed by modelling offers a plausible explanation for the observed constancy of oscillation frequency and decay across GABApre-depleted mice. While the linear nature of the model explains primary oscillatory peak frequency, it leaves open the origins of sub-harmonic peaks. During periodic cycles of movement, sensory input occasionally fails to drive supra-threshold motor neuron activation[43], suggesting that sub-harmonics could have their basis in the intermittent skipping of one or more feedback-driven oscillatory cycles. Taken together, a simple model of sensory gain control provides theoretical support for the idea that GABApre neurons have a crucial role in ensuring smooth movement.

We note that the impact of manipulating presynaptic inhibition appears highly selective. Forelimb oscillations emerge after GABApre ablation despite the persistence of postsynaptic inhibition. Parallel theoretical and experimental analyses show that the subtractive nature of postsynaptic inhibition is far less effective in gain scaling than the divisive normalization achieved by presynaptic inhibition (Supplementary Discussion)[12–14]. Moreover, experimental inactivation of other defined spinal interneuron subtypes, both excitatory and inhibitory, impairs motor performance without obvious limb oscillation[31,44–46], implying a specialized contribution of presynaptic inhibition to smooth movement. Our findings also emphasize the modular nature of skilled reach, in that oscillations after GABApre ablation are evident only during the forelimb reach phase, leaving the initiation and later grab phases unscathed. Such modular selectivity implies contextual recruitment of presynaptic inhibition during movement, permitting flexibility in the scaling of sensory gain (Supplementary Discussion)[13]. The precision of recruitment of GABApre neurons by sensory and descending pathways[4–6,47–50], when coupled with the diversity of sensory neurons influenced by presynaptic inhibition[6], hints at the existence of many GABApre neuronal subtypes, each devoted to gain control across discrete sensory feedback channels.

METHODS SUMMARY

Procedures performed in this study were conducted according to US National Institutes of Health guidelines for animal research and were approved by the Institutional Animal Care and Use Committee of Columbia University. See Methods for further details.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 December 2013; accepted 24 March 2014.

1. Windhorst, U. Muscle proprioceptive feedback and spinal networks. *Brain Res. Bull.* **73,** 155–202 (2007).
2. Arber, S. Motor circuits in action: specification, connectivity, and function. *Neuron* **74,** 975–989 (2012).
3. Frank, K. & Fuortes, M. Presynaptic and postsynaptic inhibition of monosynaptic reflexes. *Fed. Proc.* **16,** 49–50 (1957).
4. Eccles, J. C., Eccles, R. M. & Magni, F. Central inhibitory action attributable to presynaptic depolarization produced by muscle afferent volleys. *J. Physiol. (Lond.)* **159,** 147–166 (1961).
5. Lundberg, A. Supraspinal control of transmission in reflex paths to motoneurones and primary afferents. *Prog. Brain Res.* **12,** 197–221 (1964).
6. Rudomin, P. & Schmidt, R. F. Presynaptic inhibition in the vertebrate spinal cord revisited. *Exp. Brain Res.* **129,** 1–37 (1999).

7. Betley, J. N. *et al.* Stringent specificity in the construction of a GABAergic presynaptic inhibitory circuit. *Cell* **139,** 161–174 (2009).
8. Conradi, S. Ultrastructure of dorsal root boutons on lumbosacral motoneurons of the adult cat, as revealed by dorsal root section. *Acta Physiol. Scand.* **332** (Suppl.), 85–115 (1969).
9. Rossignol, S., Dubuc, R. & Gossard, J.-P. Dynamic sensorimotor interactions in locomotion. *Physiol. Rev.* **86,** 89–154 (2006).
10. Akay, T., Tourtellotte, W. G., Arber, S. & Jessell, T. M. Locomotor pattern degrades in absence of proprioceptive feedback: a genetic, physiological, and behavioral analysis. (Society for Neuroscience, 2013).
11. Stein, R. B. & Oğuztöreli, M. N. Tremor and other oscillations in neuromuscular systems. *Biol. Cybern.* **22,** 147–157 (1976).
12. Capaday, C. & Stein, R. B. A method for simulating the reflex output of a motoneuron pool. *J. Neurosci. Methods* **21,** 91–104 (1987).
13. Capaday, C. & Stein, R. B. Difference in the amplitude of the human soleus H reflex during walking and running. *J. Physiol. (Lond.)* **392,** 513–522 (1987).
14. Capaday, C. A re-examination of the possibility of controlling the firing rate gain of neurons by balancing excitatory and inhibitory conductances. *Exp. Brain Res.* **143,** 67–77 (2002).
15. Hughes, D. I. *et al.* P boutons in lamina IX of the rodent spinal cord express high levels of glutamic acid decarboxylase-65 and originate from cells in deep medial dorsal horn. *Proc. Natl Acad. Sci. USA* **102,** 9038–9043 (2005).
16. Soghomonian, J.-J. & Martin, D. L. Two isoforms of glutamate decarboxylase: why? *Trends Pharmacol. Sci.* **19,** 500–505 (1998).
17. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71,** 995–1013 (2011).
18. Sonner, P. M. & Ladle, D. R. Early postnatal development of GABAergic presynaptic inhibition of Ia proprioceptive afferent connections in mouse spinal cord. *J. Neurophysiol.* **109,** 2118–2128 (2013).
19. Behar, T. *et al.* Many spinal cord cells transiently express low molecular weight forms of glutamic acid decarboxylase during embryonic development. *Brain Res. Dev. Brain Res.* **72,** 203–218 (1993).
20. Zhang, F., Wang, L.-P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods* **3,** 785–792 (2006).
21. Willis, W. D. Dorsal root potentials and dorsal root reflexes: a double-edged sword. *Exp. Brain Res.* **124,** 395–421 (1999).
22. Barker, J. L. & Nicoll, R. A. Gamma-aminobutyric acid: role in primary afferent depolarization. *Science* **176,** 1043–1045 (1972).
23. Doyle, M. W. & Andresen, M. C. Reliability of monosynaptic sensory transmission in brain stem neurons *in vitro*. *J. Neurophysiol.* **85,** 2213–2223 (2001).
24. Eccles, J. C., Schmidt, R. & Willis, W. D. Pharmacological studies on presynaptic inhibition. *J. Physiol. (Lond.)* **168,** 500–530 (1963).
25. Jonas, P., Bischofberger, J. & Sandkühler, J. Corelease of two fast neurotransmitters at a central synapse. *Science* **281,** 419–424 (1998).
26. Kuno, M. Mechanism of facilitation and depression of the excitatory synaptic potential in spinal motoneurones. *J. Physiol. (Lond.)* **175,** 100–112 (1964).
27. Clements, J. D., Forsythe, I. D. & Redman, S. J. Presynaptic inhibition of synaptic potentials evoked in cat spinal motoneurones by impulses in single group Ia axons. *J. Physiol. (Lond.)* **383,** 153–169 (1987).
28. Li, Y. & Burke, R. E. Short-term synaptic depression in the neonatal mouse spinal cord: effects of calcium and temperature. *J. Neurophysiol.* **85,** 2047–2062 (2001).
29. Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. Synaptic depression and cortical gain control. *Science* **275,** 221–224 (1997).
30. Seki, K., Perlmutter, S. I. & Fetz, E. E. Sensory input to primate spinal cord is presynaptically inhibited during voluntary movement. *Nature Neurosci.* **6,** 1309–1316 (2003).
31. Azim, E., Jiang, J., Alstermark, B. & Jessell, T. M. Skilled reaching relies on a V2a propriospinal internal copy circuit. *Nature* **508,** 357–363 (2014).
32. Whishaw, I. Q. An endpoint, descriptive, and kinematic comparison of skilled reaching in mice (*Mus musculus*) with rats (*Rattus norvegicus*). *Behav. Brain Res.* **78,** 101–111 (1996).
33. Farr, T. D., Liu, L., Colwell, K. L., Whishaw, I. Q. & Metz, G. A. Bilateral alteration in stepping pattern after unilateral motor cortex injury: a new test strategy for analysis of skilled limb movements in neurological mouse models. *J. Neurosci. Methods* **153,** 104–113 (2006).
34. Prochazka, A. Quantifying proprioception. *Prog. Brain Res.* **123,** 133–142 (1999).
35. Laughlin, S. B. Matching coding, circuits, cells, and molecules to signals: general principles of retinal design in the fly's eye. *Prog. Retin. Eye Res.* **13,** 165–196 (1994).
36. Brenner, N., Bialek, W. & de Ruyter van Steveninck, R. Adaptive rescaling maximizes information transmission. *Neuron* **26,** 695–702 (2000).
37. Eccles, J. C., Kostyuk, P. G. & Schmidt, R. F. Central pathways responsible for depolarization of primary afferent fibres. *J. Physiol. (Lond.)* **161,** 237–257 (1962).
38. Tian, N. *et al.* The role of the synthetic enzyme GAD65 in the control of neuronal gamma-aminobutyric acid release. *Proc. Natl Acad. Sci. USA* **96,** 12911–12916 (1999).
39. Churchland, M. M. & Lisberger, S. G. Experimental and computational analysis of monkey smooth pursuit eye movements. *J. Neurophysiol.* **86,** 741–759 (2001).
40. Baker, S. N. Oscillatory interactions between sensorimotor cortex and the periphery. *Curr. Opin. Neurobiol.* **17,** 649–655 (2007).
41. Llinás, R. R. Inferior olive oscillation as the temporal basis for motricity and oscillatory reset as the basis for motor error correction. *Neuroscience* **162,** 797–804 (2009).
42. Whitmer, D. *et al.* High frequency deep brain stimulation attenuates subthalamic and cortical rhythms in Parkinson's disease. *Front. Hum. Neurosci.* **6,** 155 (2012).
43. Evans, C. M., Fellows, S. J., Rack, P. M., Ross, H. F. & Walters, D. K. Response of the normal human ankle joint to imposed sinusoidal movements. *J. Physiol. (Lond.)* **344,** 483–502 (1983).
44. Gosgnach, S. *et al.* V1 spinal neurons regulate the speed of vertebrate locomotor outputs. *Nature* **440,** 215–219 (2006).
45. Akay, T., Fouad, K. & Pearson, K. G. New technique for drug application to the spinal cord of walking mice. *J. Neurosci. Methods* **171,** 39–47 (2008).
46. Bui, T. V. *et al.* Circuits for grasping: spinal dI3 interneurons mediate cutaneous control of motor behavior. *Neuron* **78,** 191–204 (2013).
47. Rudomin, P., Jimenez, I., Solodkin, M. & Duenas, S. Sites of action of segmental and descending control of transmission on pathways mediating PAD of Ia- and Ib-afferent fibers in cat spinal cord. *J. Neurophysiol.* **50,** 743–769 (1983).
48. Lomelí, J., Quevedo, J., Linares, P. & Rudomin, P. Local control of information flow in segmental and ascending collaterals of single afferents. *Nature* **395,** 600–604 (1998).
49. Hultborn, H., Meunier, S., Pierrot-Deseilligny, E. & Shindo, M. Changes in presynaptic inhibition of Ia fibres at the onset of voluntary contraction in man. *J. Physiol. (Lond.)* **389,** 757–772 (1987).
50. Nielsen, J. & Kagamihara, Y. The regulation of presynaptic inhibition during co-contraction of antagonistic muscles in man. *J. Physiol. (Lond.)* **464,** 575–593 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to K. Miao for assistance with mouse genotyping; M. Mendelsohn, N. Zabello and S. Patruni for animal care; and B. Han, K. MacArthur, S. Morton and I. Schieren for technical assistance. We thank G.Z. Mentis for the custom built recording stage used in the electrophysiology experiments; K. Deisseroth for the adeno-associated viral *FLEX-ChR2-YFP* plasmid construct; and M. Churchland for providing a low-pass filter. We are grateful to C.E. Schoonover for sustained encouragement, engagement and discussion; B. Babadi for assistance in implementing the short-term depression model; the instructors of the Ion Channel Physiology course at Cold Spring Harbor Laboratory for guidance; and S. Siegelbaum and A. Losonczy for advice on electrophysiology. We thank R. Axel, S. Druckmann, C. Jahr, A. Karpova, A. Miri, K. Seki and S. Siegelbaum for valuable discussion and comments on the manuscript. E.A. is a Howard Hughes Medical Institute Fellow of the Helen Hay Whitney Foundation; J.Z.H. was supported by NIH grant MH078844; L.F.A. was supported by NIH grant MH093338 and by the Harold and Leila Y. Mathers, Gatsby and Swartz Foundations; T.M.J. was supported by NIH grant NS033245, the Harold and Leila Y. Mathers Foundation and Project A.L.S., and is an investigator of the Howard Hughes Medical Institute.

Author Contributions A.F. and T.M.J. devised the project. A.F., E.A. and T.M.J. designed the experiments and analysed the data. A.F. performed the anatomical and electrophysiological experiments. E.A. developed the behavioural assay and with A.F. and K.R.C. performed the behavioural experiments. L.F.A. developed and with A.F. implemented the computational models. Z.J.H. generated the *Gad2*^{Cre} mouse. A.F., E.A., L.F.A and T.M.J. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M.J. (tmj1@columbia.edu) or A.F. (af2243@columbia.edu).

METHODS

Mice. *Gad2::IRES::Cre*[17] (*Gad2^{Cre}*) and *Hb9::GFP*[51] (*Hb9^{GFP}*) mice were all on a C57BL/6 background. Animals were housed individually with light on a 12:12 h cycle. For electrophysiology experiments, which required neonates, *Gad2$^{Cre/Cre}$* homozygous mice were paired for breeding. Procedures performed in this study were conducted according to US National Institutes of Health guidelines for animal research and were approved by the Institutional Animal Care and Use Committee of Columbia University.

Viral vectors. To mark Cre-recombinase expressing neurons in *Gad2^{Cre}* mice for photoactivation or anatomical tracing, we used an adeno-associated viral (AAV) construct in which an inverted *hChR2(H134R)-EYFP* sequence was flanked by two pairs of heterotypic antiparallel *loxP* recombination sites (K. Deisseroth; Addgene plasmid 20298). *AAV-FLEX-hChR2-YFP* (*FLEX-ChR2-YFP*) was packaged and pseudotyped with an AAV1 serotype to a titre of 4×10^{12} viral particles per ml (UNC Gene Therapy Center). For anatomical analysis and photoactivation experiments we performed *FLEX-ChR2-YFP* injection at p0–3 into lumbar spinal cord. For anatomical tracing of GABApre neurons we performed *FLEX-ChR2-YFP* injection at p21–30 into lumbar spinal cord and ~p78 into cervical spinal cord.

To mark *Gad2^{Cre}* neurons for acute ablation we generated a Cre-dependent *AAV-FLEX-DTR-GFP* (*FLEX-DTR-GFP*), in which a simian *DTR (HBEGF)* sequence was fused to *GFP*. Details for the construction of the *FLEX-DTR-GFP* construct are described elsewhere[31]. Briefly, the insert was flanked by two pairs of heterotypic, antiparallel *loxP* recombination sites, downstream of a *CAG* promoter and upstream of a woodchuck hepatitis virus post-transcriptional regulatory element (*WPRE*) sequence. After confirming Cre-mediated expression specificity, the virus was packaged and pseudotyped with an AAV1 serotype to a titre of 5×10^{12} viral particles per ml (UNC Gene Therapy Center). To drive DTR expression in *Gad2^{Cre}* neurons we performed *FLEX-DTR-GFP* injection at p56–84 into cervical spinal cord.

Lumbar viral injections for photoactivation and anatomical experiments. *Gad2^{Cre}* pregnant females were housed individually ~5 days before giving birth and given additional bedding materials to encourage nest building.

Before removing pups from the nest, the mother was placed in a separate cage so as not to be disturbed. All pups were removed from the nest and placed on a heating pad throughout the injection procedure. For injections, individual pups were anaesthetized with isoflurane. To gently stretch the vertebral column the feet and head of the pup were taped to a paper towel placed on a heating pad. A small, longitudinal incision was made in the skin between the hips and shoulder, directly over the spinal cord. Connective tissue covering the vertebral column was gently removed using fine scissors and the skin flaps were retracted to expose the vertebrae. The posterior-most rib was identified and used to locate the L1 vertebral lamina.

Injections were performed using a Nanoject II (Drummond Scientific Company) placed in a stereotaxic arm (Narishege) with glass capillaries pulled using a P-30 vertical puller (Sutter) and then cut under a dissection stereomicroscope with fine scissors to ensure a sharp point. The Nanoject II was set to deliver 23 nl of solution per injection on the slow setting.

We targeted our injections to the medial portion of the spinal cord where GABApre neurons reside. We positioned the tip of the capillary 50 μm to the right of midline and in between the laminae (at this young age, it is possible to insert the fine capillary needle directly in between the laminae into the spinal cord). We then penetrated 1,200 μm into the tissue and injected 23 nl of solution every 100 μm while retracting. Following each injection we waited 10 s to allow diffusion of the solution. At the top of each injection track we waited for 1 min to allow for diffusion of the solution and to prevent virus from exiting the cord with the capillary. Injections were performed at the three laminae above and below the L1 vertebra, permitting delivery of virus throughout the lumbar spinal cord. The incision was then closed using four inside-out sutures, and the pups were allowed to recover on the heating pad until awake, at which point they were returned to the nest. Only several minutes after returning the pups to the nest was the mother reintroduced to the home cage.

To characterize the population of *Gad2*-expressing neurons at later ages, we injected *FLEX-ChR2-YFP* into lumbar spinal cord at p21–30. Mice were anaesthetized with isoflurane and the fur covering the lumbar vertebral segments was removed. To prevent dryness of the eyes, eye lubricant (Puralube Vet Ointment; Dechra) was applied. Similar to neonate injections, the animal was securely positioned by gently taping the hindlimbs, forelimbs, shoulders and hips, and an incision was made in the skin to expose lumbar and caudal thoracic spinal segments. It was not necessary to remove any muscle above the spinal vertebrae, although connective tissue covering the segments was typically removed to allow penetration of the glass capillary. Since the spinal cord is larger at these older ages, injections were made 1,500 μm below the surface, every 100 μm, following the same procedure as at p0–3. Three segments rostral and caudal to the L1 vertebra were injected, resulting in AAV delivery to the full extent of the lumbar spinal cord.

Cervical viral injections for anatomical and ablation experiments. A detailed description of the cervical injection procedure can be found elsewhere[31]. Briefly, for cervical spinal cord AAV injections, mice were anaesthetized with 0.01 ml g^{-1} body weight of 2.5% tribromoethanol (Sigma-Aldrich) via intraperitoneal injection and maintained as needed. Eye lubricant was applied and the mice were placed in a stereotaxic frame (David Kopf Instruments). The head of the mouse was tilted forward and the tail was gently elevated and pulled back using a spinal vertebrae clamp (David Kopf Instruments). After making an incision in the skin covering the cervical spinal cord, the cervical laminae were exposed using forceps and a delicate bone scraper (Fine Science Tools). The laminae were cleaned with absorption spears (Fine Science Tools) and gently separated via retraction of the large spiny process of the second thoracic vertebra using a small alligator clip held in place by a spinal vertebrae clamp. After removal of the cervical dura mater using fine forceps, *FLEX-ChR2-YFP* virus (for anatomical tracing) or *FLEX-DTR-GFP* virus (for ablation experiments) was delivered to cervical spinal cord using a Nanoject II. Injections were restricted to the right side of the C3–T1 segments, with three injection traces per segment, 23 nl per injection, spanning the dorso-ventral extent of the cord.

For ablation experiments, approximately 14–21 days following *FLEX-DTR-GFP* injection, mice were administered 400 ng of diphtheria toxin (Sigma-Aldrich) in sterile phosphate buffered saline via intraperitoneal injection. For analysis of ablation efficiency, mice were perfused one week after DT administration and tissue was processed for immunohistochemistry.

Electrophysiology, dissection. To obtain whole-cell patch-clamp recordings from motor neurons with intact sensory input and interneuron circuitry, we developed an *in vitro* semi-intact spinal cord preparation that permitted recordings from spinal motor neurons under visual guidance. Mice aged p9–14 were anaesthetized with tribromoethanol (2.5% solution, 30 μl per gram body weight) and transcardially perfused using ice cold dissection artificial cerebral spinal fluid (dACSF) containing 234 mM sucrose, 3.6 mM KCl, 1.2 mM $NaH_2PO_4.H_2O$, 25 mM $NaHCO_3$, 11 mM D-glucose, 0.5 mM $CaCl_2$, 4 mM $MgCl_2$, 0.4 mM ascorbic acid, 2 mM pyruvate, 20 μM APV and 5 mM kynurenic acid, and equilibrated with 95% O_2 and 5% CO_2. We found that effective transcardial perfusion was critical to obtain healthy motor neurons at these ages.

The vertebral column was then extracted and placed in a dissection dish superfused with oxygenated dACSF and cooled to ~2 °C using a custom made, chilled brass chamber (G. Johnson, Columbia University). The spinal cord and rib cage was pinned ventral side up in a Sylgard 184 Silicone Elastomer (Dow Corning)-coated dissection dish and the corpora were removed. The thoracic and sacral portions of the spinal cord were removed and a longitudinal incision was made in the dura mater along the ventral surface of the lumbar spinal cord using fine Vannas scissors (FST). Using no. 55 forceps (FST) the dura was then peeled off of the ventral surface on the right side of the spinal cord, taking care not to damage the dorsal roots. The cord was then hemisected by passing a bevelled crescent scalpel (Sharptome) through the midline. The dorsal roots on the right side were then cut and the right half of the cord gently removed from the vertebral column. The left hemicord was used in some experiments for ventral and dorsal root recordings, coupled with photoillumination of the exposed midline, following standard methods[18,52].

To expose motor neurons for whole-cell recordings we removed the ventral-lateral white matter. To prepare the tissue for cutting, we embedded the cord in 1.5% low melting point (LMP) agar made fresh with dACSF and incubated at 34 °C during the dissection. We placed the hemicord in a $15 \times 15 \times 5$ mm disposable base mould (Fisher) and carefully positioned it so that the ventral surface was facing down, towards the base of the mould. We then quickly replaced the dACSF with 1.5% LMP agar, siphoning off the dACSF with a Pasteur pipette and applying the warm, LMP agar with a plastic transfer pipette. Once the cord was in the LMP agar, but before the agar had hardened, we carefully lifted the dorsal roots away from the cutting plane, making sure that the cord was positioned so that the dorsal–ventral axis was perpendicular to the bottom of the mould. We then placed the mould on a cold plate and covered the surface with a chilled glass coverslip to speed setting of the agar. It is important to place the mould on the cold plate only once the cord and roots are in position, as the agar sets quickly making further manipulation of the tissue difficult. The mould and coverslip were then immediately submerged in ice cold, oxygenated dACSF and the coverslip was removed once the agar was sufficiently hardened (after about ten seconds). Submerged in the solution, the agar block containing the hemicord was then extracted from the mould using a small metal spatula.

The agar block was glued using Loctite 404 cyanoacrylate glue to an angled block cut from 4% agar (~25° angle) which itself had been glued to the chilled vibratome stage. The stage was then quickly placed in ice cold, oxygenated dACSF in the vibratome bath. A vibratome (Leica VT1200) was used to remove the ventral-lateral white matter; the vibratome blade was positioned at the surface of the spinal cord, then lowered 50 μm below the surface and used to cut a sliver of white matter off of the hemicord. Cuts were made, typically between one and three, until a fine strip of

grey matter was visible through a stereo dissection scope. The cut hemicord cord was then carefully removed from the 1.5% agar and placed in recovery solution. The hemicord was then allowed to recover for 30 min at 34 °C before being placed in the recording chamber (Warner) where it was superfused with recording ACSF containing 125 mM NaCl, 26 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 2.5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 0.4 mM ascorbic acid, 2 mM pyruvate and 26 mM D-glucose, and equilibrated with 95% O_2 and 5% CO_2. The recovery solution was a 50:50 mixture of ACSF and dASCF.

Electrophysiology recordings. The cord was placed in a laminar flow chamber (Warner) with the exposed ventral surface facing up and secured with a harp strung with fine, loose, nylon fibres. Dorsal roots L3, L4 and L5 were positioned within polyethylene suction electrodes for nerve stimulation. The suction electrodes were held using a custom made holder and miniature manipulators (G.Z. Mentis, Columbia University). A stimulus isolator (AMPI) was used to stimulate dorsal roots with 100 µs current pulses. Dorsal roots were stimulated at 1.2-fold threshold, with threshold defined as the stimulation intensity that generated a postsynaptic response in roughly 60% of trials[52]. Stimulation intensities were typically between 2 and 10 µA, depending on the size of the root and the seal of the suction electrode.

Whole-cell recordings were obtained using 2.5–3.5 MΩ pipettes pulled using a P-2000 puller (Sutter). An Olympus BX-51 microscope with an infrared DIC filter set was used to visualize motor neurons. A Multiclamp 700B amplifier and Digidata 1440A (Molecular Devices) were used to record and transfer electrical potentials to a desktop computer (Dell). Data were digitized at 20–50 kHz and a 10 kHz low-pass filter was used before digitization. Motor neurons were identified based on their ventral-lateral position, their large size and distinct dendritic morphology, their electrophysiological properties and their expression of GFP using *Hb9*^{GFP} transgenic mice. Motor neurons typically had a resting membrane potential between −68 and −74 mV. Cells with a resting membrane potential more depolarized than −60 mV were discarded.

Voltage clamp experiments were performed using a caesium-methanesulphonate-based internal solution containing 135 mM $CsMeSO_3$, 5 mM $CsCl_2$, 2 mM NaCl, 10 mM HEPES, 0.1 mM EGTA, 5 mM MgATP, 0.4 mM Na_2GTP, 10 mM phosphocreatine and 5 mM QX-314 chloride. Motor neurons were held between −60 and −70 mV and recordings in which the series resistance exceeded 20 MΩ or increased by greater than 20% during a recording were excluded. Because of the large size and low input resistance of the motor neurons, as well as the magnitude of the synaptic currents, series resistance was monitored but not compensated. Current clamp recordings were performed using a potassium-gluconate-based internal solution with 130 mM potassium gluconate, 5 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES and 4 mM Na_2ATP. All experiments in which sensory-evoked synaptic currents were isolated were performed using 20 µM APV (Tocris). Internal solutions were titrated to a pH of 7.4 with potassium hydroxide for the potassium-gluconate-based internal solution and caesium hydroxide for the caesium-methanesulphonate-based internal solution. Osmolarity was between 290 and 300 mOsm and was measured using a vapour pressure osmometer (Vapro).

To avoid potential confounds of voltage-gated channel activation during synaptic excitation, we applied internal blockage of voltage-gated potassium channels (via caesium) and sodium channels (via QX-314). However, blocking sodium and potassium channels increases motor neuron input resistance that, together with the high capacitance of spinal motor neurons due to their extreme size, results in prolonged membrane time constants. Under these conditions excitatory postsynaptic potentials (EPSPs), but not EPSCs, have an extended decay time. Thus, except for our mean variance analysis experiments (see below), we typically used voltage clamp to isolate monosynaptic sensory input to motor neurons.

Extracellular recordings from dorsal roots and ventral roots were obtained using the same suction electrodes used for nerve stimulation. Signals were amplified using a Cyberamp 320 extracellular amplifier (Molecular Devices) and were low-pass filtered at 10,000 kHz and in some cases high-pass filtered at 0.1 or 1 Hz.

Drugs used: the GABA-A antagonist SR 95531 hydrobromide (gabazine, 2 µM), the GABA-B antagonist CGP 54626 (CGP, 2.5 µM), the glycine receptor antagonist strychnine (5 µM), the NMDA receptor antagonist APV (20 µM) and the AMPA receptor antagonist NBQX (10 µM), all from Tocris.

Photostimulation was performed using a 473 nm diode laser (CrystaLaser) controlled by 5 V transistor–transistor logic (TTL) pulses of varying duration. Laser strength was set to be suprathreshold for generation of IPSCs in motor neurons and pulse length was typically 1 ms, although in some experiments 5 ms pulses were used. For calibration of spiking responses of *Gad2*^{Cre} neurons the laser pulse width was 1 ms. Recordings from ChR2$^+$ *Gad2*^{Cre} neurons were obtained using a sagittal hemisected preparation[53] that permits targeted recordings from neurons residing near the midline. The central canal was identified and ChR2–YFP$^+$ neurons residing within 200 µm dorsal and 100 µm lateral of the central canal were targeted. In some preparations we noted that IPSCs followed light pulse frequencies greater than 50 Hz, presumably due to high levels of ChR2 expression. We therefore used 100 Hz light pulses in some experiments to ensure maximal *Gad2*^{Cre} neuron activation.

All electrophysiology data were analysed in MATLAB (Mathworks). Sensory-evoked EPSC onset was calculated by identifying the time point at which the derivative of the EPSC waveform crossed a noise threshold (twice the variance of the derivative waveform in the 50 ms preceding stimulation) for five consecutive time bins. Monosynaptic inputs were identified based on the low variance of their trial-to-trial onset latency ($cv_{onset} < 0.05$)[23]. Coefficient of variation was calculated as the standard deviation divided by the mean throughout. The onset of IPSCs, as well as ventral root and dorsal root spikes, were estimated using a similar thresholding method. Instead of using the derivative trace, however, which was at times too noisy, onset was defined as the point where the current or voltage waveform crossed a hand-picked threshold for five consecutive points. The estimated onset times were then checked by eye for each trace.

The IPSC index ($IPSC_{index}$, Fig. 3d) was calculated as follows, where $IPSC_{\lambda,\Phi}$ is the mean amplitude of light-evoked IPSCs in the presence of drug (either strychnine or SR 95531 and CGP 54626) and $IPSC_\lambda$ is the mean amplitude of light-evoked IPSCs in the absence of drug:

$$IPSC_{index} = \frac{IPSC_{\lambda,\Phi}}{IPSC_\lambda} \tag{1}$$

The EPSC index ($EPSC_{index}$, Fig. 3g) was calculated as follows, where $EPSC_{\lambda,\Phi}$ is the mean sensory-evoked EPSC amplitude following *Gad2*^{Cre}-neuron photoactivation in the presence of drug, $EPSC_\Phi$ is the mean sensory-evoked EPSC amplitude without *Gad2*^{Cre} neuron photoactivation in the presence of drug, $EPSC_\lambda$ is the mean sensory-evoked EPSC amplitude following *Gad2*^{Cre} neuron photoactivation without drug and EPSC is the mean sensory-evoked EPSC amplitude with neither *Gad2*^{Cre} neuron photoactivation nor drug present:

$$EPSC_{index} = \frac{1 - \dfrac{EPSC_{\lambda,\Phi}}{EPSC_\Phi}}{1 - \dfrac{EPSC_\lambda}{EPSC}} \tag{2}$$

The value of the IPSC index is 1 when the drug has no effect and 0 when the drug completely blocks the IPSC. The value of the EPSC index is 1 when the drug has no effect on *Gad2*^{Cre} neuron-mediated suppression of sensory-evoked EPSCs and 0 when the drug abolishes the suppression of sensory-evoked EPSCs achieved by *Gad2*^{Cre} neuron photoactivation.

For intracellular recording data *n* values indicate neurons recorded. For extracellular root recordings *n* values indicate preparations.

Vesicle depletion model. The vesicle depletion model was adapted from previously described simulations[29,54] and was implemented in MATLAB. The model has three parameters: the time constant of facilitation, the time constant of depression and release probability. The facilitation time constant was not found to influence vesicle depletion at frequencies ≤25 Hz, and so was set to 10 ms. The depression time constant was set so that normalized synaptic conductances generated by the model matched by eye those from control (no photostimulation) electrophysiological data. We found that depression time constants between 125 ms and 975 ms with a mean of 437 ± 305 ms s.d. captured the depression kinetics we observed. To determine the release probability time constant, the average amplitudes of consecutive experimentally-derived EPSCs were estimated. EPSC amplitudes were then generated using the vesicle depletion model and were fit with single exponential functions, which were then superimposed upon normalized EPSC amplitudes obtained from actual recordings. Changing the release probability parameter of the model, and only the release probability parameter, was sufficient to capture light-evoked changes in the rate of synaptic depression.

Mean variance analysis. The mean variance analysis used was described previously[55]. Data were collected in current clamp to avoid measurement errors introduced by changes in series resistance over long recording sessions (Extended Data Fig. 6). Each calcium concentration was visited twice for a given cell, and EPSP amplitudes used for analysis were taken from two sets of 25 consecutive dorsal root stimuli once the amplitude had stabilized following introduction of a new calcium concentration. Quadratic functions were fit using standard curve fitting routines in MATLAB. The cv_q parameter (coefficient of variation in EPSP amplitude from a single release site) in the mean variance analysis was set to 0.3, following established values[55]. Changes (± 10%) to this parameter did not markedly affect the release probability estimates.

Kinematic analysis of reaching. The basic training paradigm and reaching box design were largely based on established methods[32], and were adapted to 3D kinematic analysis in the mouse, as described[31].

Briefly, adult mice (~12–20 weeks old) were food deprived to ~80–90% of their original body weight and habituated to target food pellets (20 mg, 3 mm diameter; Bio-Serv). Approximately 3 g of pellets were provided at the end of each

day to maintain food deprivation. Following habituation, mice were placed into a clear acrylic box with a narrow opening[31]. Food pellets were placed 1 cm away from the opening to the left of centre, to encourage use of the right arm. Mice were acclimated and trained to reach for three to four days, and the mice that were readily reaching with the right paw were selected for experiments (~50–75% of mice). Mice were then trained for two weeks for 20 min or 20 successful reaches each day, whichever occurred first.

For three dimensional quantification of reaching kinematics, mice were placed in a reaching box with a modified pellet tray[31] in view of two high-speed, high-resolution monochrome cameras (HiSpec 1; 2 GB memory; Fastec) with 50 mm *f*/1.4 manual iris and focus lenses (C-Mount, 1/2 inch CCD; Fujinon) placed ~80° apart. An infrared LED light source (Clover Electronics) was mounted on top of each camera, and each lens was equipped with an infrared long-pass filter (Midwest Optical Systems). Cameras were synced to each other, calibrated in space using MaxTRAQ 3D software (Innovision Systems) and set to 500 frames per second with a resolution of 1,280 × 1,024. A black background was used to increase contrast.

Mice were briefly anaesthetized with isoflurane, and a 1.5 mm reflective hemisphere marker (B & L Engineering) was attached to the back of the right paw with adhesive. HiSpec camera control software was used to record reach trials, using an external trigger in ring mode. MaxTRAQ 2D software (Innovision Systems) was used to automatically track the marker and the pellet, with trials beginning as the paw left the ground and ending when the paw contacted the pellet, or when a full reach and prehension motion was completed. Tracked files were imported into MaxTRAQ 3D software to compute the coordinates of the marker and pellet in three dimensions.

Data were imported into MATLAB, low-pass filtered, and the starting points of reach trials were normalized by discarding reaches that started at a z distance (between paw and target) below 13 mm, and removing data frames for the remaining reaches above the 13 mm threshold (post-normalization: 112 DTR reaches, 75 control reaches; see ref. 31). Analyses of early reach and late grab phases were performed before and after a distance-to-pellet threshold of 11 mm, where the paw passes through the opening of the box. Mean power spectra were calculated using a fast Fourier transform of individual reach velocities. Decay times were calculated by aligning all reaches within individual mice to the maximum velocity peak of each reach, normalizing each reach to its maximum value, averaging across reaches and fitting a single exponential function to the first three estimated peaks for each mouse. The time constant of this exponential was considered the oscillatory decay time for each mouse. Power and decay analyses were performed on all reaches regardless of paw starting position, and thus non z-position normalized data were included (132 total DTR reaches). To estimate the initial and subsequent velocity peaks, the maximum velocity value was calculated for each reach in the first 50 ms, for the first peak, and the remainder of the reach, for the subsequent peak. Because all kinematic comparisons were performed within the same mouse before and after manipulation, and because tracking was automated, blinding was not performed. Analyses were performed on approximately equal numbers of male and female mice selected randomly from trained populations, and no sex-specific differences in reaching success or kinematics were identified.

Reach success was quantified over a period of 20 min or 20 hits, whichever came first, for 10–20 total days beginning after recovery from viral injection, and for 5–6 days post-DT (beginning ~2 days after DT administration). Control mice were quantified for 14–24 days pre-DT and for 8–10 days post-DT (beginning ~3–7 days after DT administration)[31].

Maximum digit abduction was calculated in MaxTRAQ 2D software by manually marking and measuring the distance between digits 2 and 4 in the frame captured by camera no. 1 in which the digits were maximally extended, as described[31]. The horizontal ladder task was performed as described[31,33]. Briefly, a high-definition camcorder (Canon Vixia HF11) was used to capture locomotion across a walkway consisting of 13 evenly-spaced rungs. An angled mirror below the walkway provided a bottom-up view of the rungs. The number of successful (hits) and unsuccessful (missteps) placements of the right and left forepaw were quantified for 10–15 full walks per mouse. The mean was calculated across all walks in individual mice pre- and post-DT. Pre-DT walks were recorded 1–2 weeks after viral injection and post-DT walks were recorded 5–7 days after DT administration.

For post-DT analysis of scratching, mice were placed in a clear box and the amount of time spent scratching the right forelimb and the rate of scratching bouts over a 10 min period were quantified. To prevent scratching-induced injury, following DT administration mice were anaesthetized with isoflurane and the claws of the right hindlimb were carefully removed with fine scissors. We observed no change in reaching behaviour due to hindlimb claw removal. To block cutaneous excitation and evaluate the effects on scratching, reaching and ladder paw placement, a topical lidocaine cream (DermaPlanet; lidocaine, benzocaine, tetracaine) was applied to the right forearm and paw. Behavioural analysis resumed ~10–15 min after lidocaine application.

Joint model. We constructed a version of the Stein–Oğuztöreli model[11] in which the angle of a single elbow-like joint (θ) is controlled by flexor (T_f) and extensor (T_e) muscle torques as follows,

$$\tau_m^2 \frac{d^2\theta}{dt^2} = T_e - T_f - k\tau_m \frac{d\theta}{dt} \tag{3}$$

The relationship between joint angle and the dimensionless flexor and extensor torque defined in equation (3) is given by Newton's second law in its angular form (where net torque is proportional to angular acceleration) combined with a damping term scaled by a dimensionless parameter k. Timescales in the model are set by a muscle time constant τ_m, so that all variables and other parameters are dimensionless. We set the damping coefficient to $k = 0.5$, but found that our results were quite insensitive to the value of k (see below and Extended Data Fig. 8).

To initiate joint extension, an excitatory drive (I_e) that peaks at 25 ms is applied to the dimensionless extensor torque in the form of a Gaussian pulse,

$$I_e = 25 \exp\left(-\frac{(t-25\ \text{ms})^2}{2(10\ \text{ms})^2}\right) \tag{4}$$

The width and magnitude of this pulse determine the magnitude of initial joint extension and were set by eye to match the time course of forepaw extension found in the reaching data. To capture sensory feedback signals generated by changes in joint angle, we included an angular velocity dependent feedback term coupled with a gain scaling factor h that determines the efficacy with which sensory feedback drives flexor and extensor torques. This term, which is purely velocity dependent, introduces a drive of magnitude $s = h\tau_m d\theta/dt$ into the muscle torque, where s is the sensory feedback term. The relationship between extensor torque, the external excitatory drive and the sensory feedback is given by

$$\tau_m^2 \frac{d^2 T_e}{dt^2} + 2\tau_m \frac{dT_e}{dt} + T_e = I_e - h\tau_m \frac{d\theta}{dt} \tag{5}$$

which corresponds to an alpha-function muscle impulse response[11]. Here, the excitatory pulse I_e serves to increase extensor torque and the sensory feedback term increases extensor torque when the joint's angular velocity is less than zero, in other words when the joint angle is decreasing so the joint is flexing. Conversely, the relationship between flexor torque and sensory feedback is given by

$$\tau_m^2 \frac{d^2 T_f}{dt^2} + 2\tau_m \frac{dT_f}{dt} + T_f = h\tau_m \frac{d\theta}{dt} \tag{6}$$

Here the sensory feedback activates the flexor torque when the angular velocity of the joint is greater than zero, which occurs when the joint angle is increasing so the joint is extending. Note that there is no external drive to the flexor torque; its activation is achieved solely by sensory feedback.

In the original Stein–Orğuztöreli model, sensory feedback depends on both a muscle length term (analogous to joint angle) and a change of muscle length term (analogous to joint angular velocity), and there is a sensory feedback delay. We found that it was not necessary to include a joint angle dependence to observe joint oscillation at high feedback gains and so, for simplicity, we did not include these terms in our model. We also included feedback delays of varying latencies (Supplementary Note 5).

The model defined by equations (3), (5) and (6) has an instability when the gain h exceeds a critical value h_c that leads to ever-increasing oscillation amplitudes. We report all gains in relation to this critical gain h_c. To generate the traces included in Fig. 6c, d, we set sensory feedback gain at two different levels. For the low-gain case, we set h to 10% of the critical gain ($h/h_c = 0.1$). Under this condition, the external extensor drive triggers joint extension, which is arrested by flexion activated by sensory feedback. The magnitude of the sensory feedback generates flexor torque of sufficient strength to arrest joint extension but of insufficient strength to generate further extensor torque. The velocity oscillation (Fig. 6c) is, in other words, near to being critically damped. In contrast, in the high-gain case, where h is 60% of the critical gain ($h/h_c = 0.6$), the sensory activated flexor torque is of sufficient strength to activate an additional extensor torque, via a pronounced and rapid decrease in joint angle. This extensor torque, in turn, triggers an increase in joint angle of sufficient velocity to trigger another flexor torque and oscillation ensues (Fig. 6d). In the high-gain case, to achieve continuous joint extension in the phase of oscillation we add a constant value to the extensor drive, setting $I_e \rightarrow I_e + 10$. We incorporated varying delays into the model by defining the contribution of the sensory feedback with a delayed angular velocity term such that for $t > t_{delay}$, $s(t) = h\tau_m \frac{d\theta}{dt}(t - t_{delay})$, where, in individual simulations, t_{delay} is taken from a set of delay values (1, 3, 5 and 10 ms) as reported in Extended Data Fig. 8. For multiple delays, we summed $\frac{d\theta}{dt}$ evaluated at different times, with equal weighting.

To simulate joint extension in the absence of subtractive scaling (representing postsynaptic inhibition) we added a constant value to the sensory feedback such

that $s = h\tau_m \frac{d\theta}{dt} + C$. The added constant C represents the added feedback that would emerge in the absence of a tonic subtractive postsynaptic inhibitory term. We set $C = 0$ for $\frac{d\theta}{dt} \leq 0$. Without this rectification, the sensory feedback generates a constant flexing torque that destabilizes the resting state of the arm. We note that even in this case there is no oscillation in the joint, but to compare its behaviour with reduced subtractive inhibition to cases without this added constant, we included the rectification step. We note that this is the only rectification in our model. In general, to keep the model linear, we did not sign-constrain either the stretch response or the muscle torques. We checked the behaviour of the model with sign-constrained stretch responses and muscle torques and note that the basic oscillatory behaviour persists at high gain, with near to critical damping at low gain.

To analyse the oscillatory behaviour of the model, we expressed the equations that describe the dynamics of the model (equations (3), (5) and (6)) as a system of five, first-order, linear differential equations:

$$\tau_m \frac{d}{dt}\begin{pmatrix} \tau_m \frac{d\theta}{dt} \\ T_e \\ T_f \\ z_e \\ z_f \end{pmatrix} = \begin{pmatrix} -k & 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ -h & 0 & 0 & -1 & 0 \\ h & 0 & 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} \tau_m \frac{d\theta}{dt} \\ T_e \\ T_f \\ z_e \\ z_f \end{pmatrix} \qquad (7)$$

where z_e and z_f are variables that permit the two second-order equations (5) and (6) to be expressed as four first-order equations. Computing the eigenvalues of the matrix describing this system of linear equations permits a description of the frequency and damping time of its dynamic modes. Three of the eigenvalues are real, corresponding to decaying modes. The eigenvalue with the least negative real part, which we call λ_1, is complex (actually it is one of a complex-conjugate pair of eigenvalues), corresponding to the dominant oscillatory mode of the model. The damping time of the oscillation is determined by the real part of λ_1 and the frequency f_{osc} of the oscillations by the imaginary part of λ_1. The damping describes the rate at which the oscillations decay. There is a level of the gain above which the damping time approaches infinity and oscillations do not decay, which defines the critical gain h_c.

To determine the appropriate values of the additional model parameters h and τ_m, we fit the oscillatory frequency and decay time of the leading eigenmode of the model to the values measured from post-DT reaching movements across mice (Extended Data Fig. 8a, b and Supplementary Note 4). The frequency of oscillation of the model is

$$f_{osc} = \frac{|\mathrm{Im}(\lambda_1)|}{2\pi\tau_m} \qquad (8)$$

where the numerator is the imaginary part of the complex eigenvalue corresponding to the model's oscillatory mode. The decay time of oscillation is,

$$\tau_{osc} = \frac{\tau_m}{|\mathrm{Re}(\lambda_1)|} \qquad (9)$$

where the denominator is the real part of the complex eigenvalue corresponding to the model's oscillatory mode. The product of these two values generates a term n_{cycle} that is independent of the muscle time constant (the muscle time constant cancels when the right side of equation (8) is multiplied by the right side of equation (9)),

$$n_{cycle} = \frac{|\mathrm{Im}(\lambda_1)|}{2\pi|\mathrm{Re}(\lambda_1)|} \qquad (10)$$

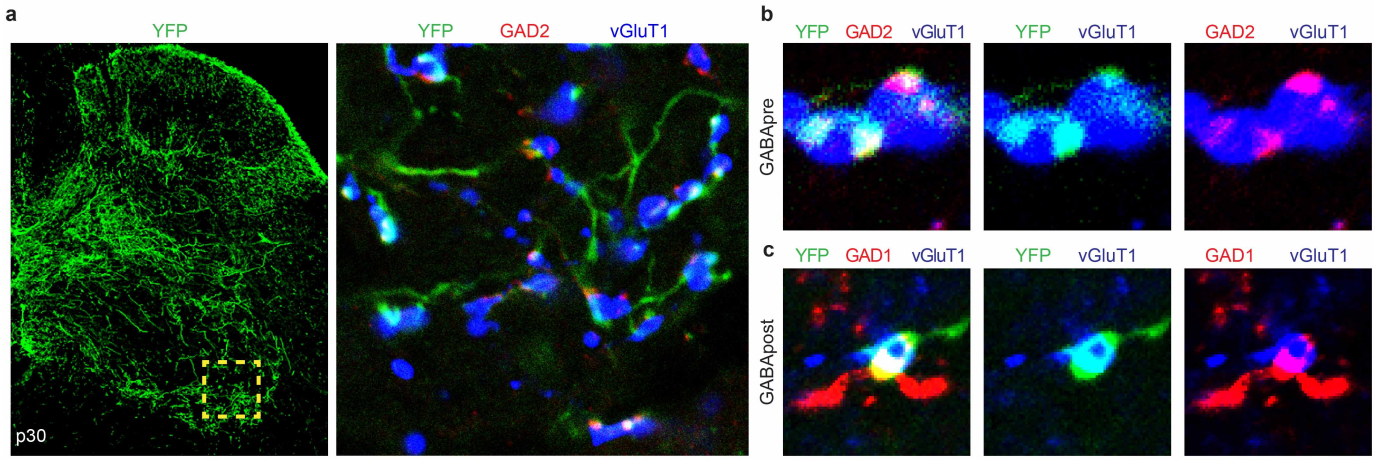
Here n_{cycle} corresponds to the number of oscillatory cycles within a decay time constant. Since we have measured the frequency and decay time of post-DT oscillation, we calculated an experimentally-derived value of n_{cycle}. We then used this value to determine which sensory gain level corresponds to the reach oscillations we observed. We also examined the extent to which the drag term k affects n_{cycle}, varying k across a 100-fold range and finding minimal change in the gain value that corresponds to our experimentally-derived n_{cycle} value (Supplementary Note 4 and Extended Data Fig. 8a). We then used this gain value to estimate the muscle time constant (Supplementary Note 4 and Extended Data Fig. 8b). We found a biologically realistic value that was within the range used in the original model[11].

Antibodies. Primary antibodies: rabbit anti-GFP (1:500, Invitrogen); sheep anti-GFP (1:1,000, AbD Serotec); rabbit anti-tdTomato (1: 1,000, Clontech; Living Colors DsRed); custom made guinea pig anti-vGluT1 (1:32,000); rabbit anti-GAD1 (1:10,000); and rabbit anti-GAD2 (1:10,000) were used as described previously[7]. Appropriate fluorophore-conjugated secondary antibodies were from the Jackson Immuno-Reseach antibody series.

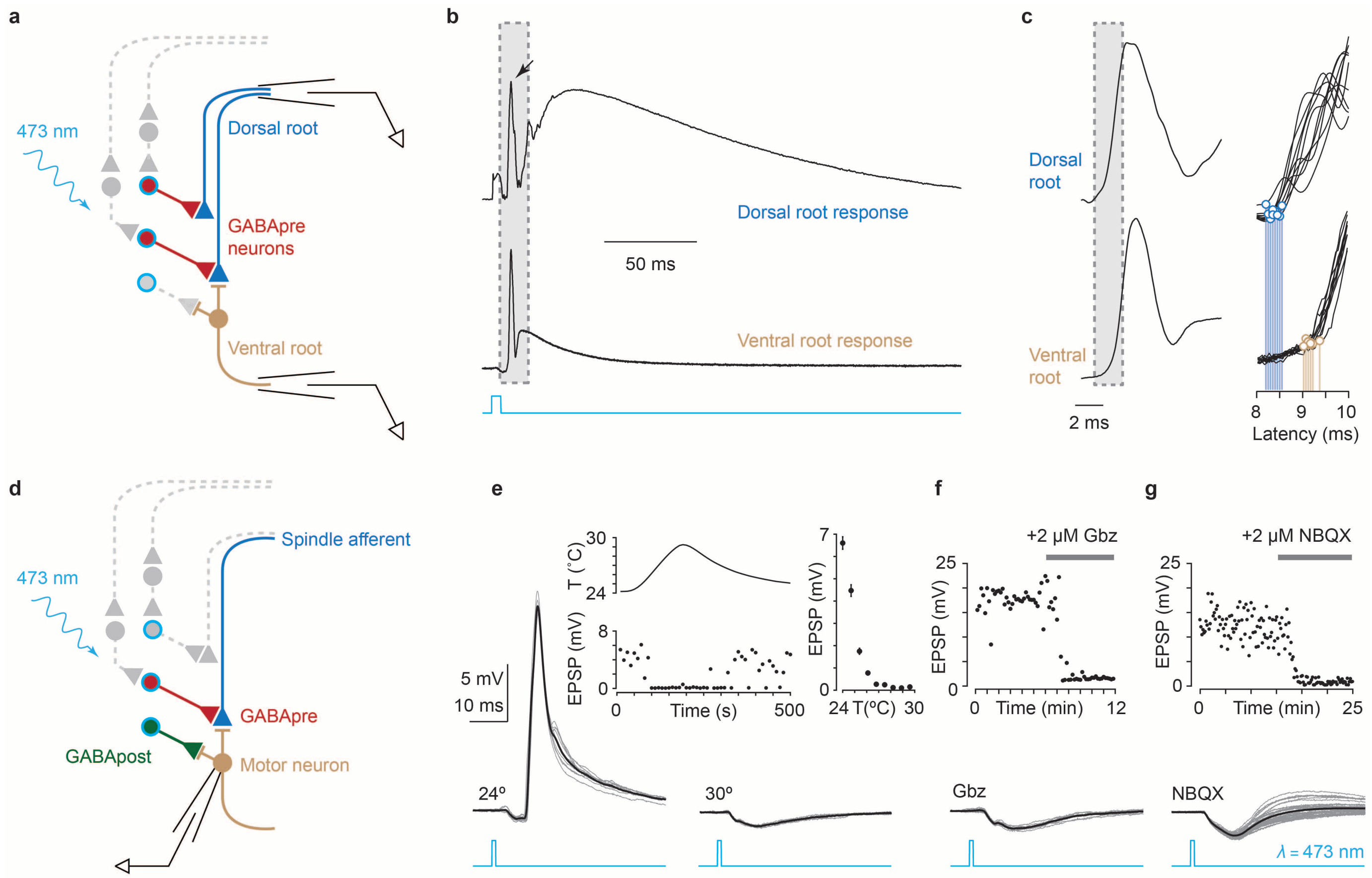
Immunohistochemistry and imaging. Mice were transcardially perfused with 4% paraformaldehyde 0.1 M phosphate buffer fixative. For cryostat spinal cord sections, tissue was post-fixed for 2 h, washed, equilibrated in 30% sucrose 0.1 M phosphate buffer solution, embedded and frozen in O.C.T. and sectioned along the axial plane at 20 μm onto glass slides. Immunohistochemistry was performed via exposure to primary antibodies (overnight at 4 °C) and fluorophore-conjugated secondary antibodies (1 h at room temperature). Sections were mounted using Fluoromount-G (SouthernBiotech) and coverslipped for imaging. Confocal images were taken with an LSM 710 microscope (Carl Zeiss).

Statistics. Results are expressed as the mean ± s.e.m. or s.d., as indicated. Electrophysiology data were analysed using two-tailed paired t tests. For behavioural experiments early tests revealed very large effects of GABApre ablation (for example see Extended Data Table 1); power analysis demonstrated that with these effect sizes relatively small sample numbers are sufficient to achieve a power of value 0.8 with an alpha value of 0.05. The large effects observed mitigate potential confounds introduced by possible violations of parametric test assumptions. Reach success and kinematic data were analysed using two-way repeated-measures ANOVA, enabling within-mouse pairing of data. Successful and unsuccessful reaches were grouped together for pre-DT conditions. There were no successful reaches following GABApre neuronal ablation in the kinematic assay. Bonferroni post hoc multiple comparisons tests were performed, as indicated. DT-treated control mice spared viral injection were analysed in parallel experiments[31] and were used in two-way repeated-measures ANOVA analysis. For velocity power spectra analysis, lidocaine scratching analysis, digit abduction and ladder walk data, two-tailed paired t tests were used. $P < 0.05$ was considered significant.

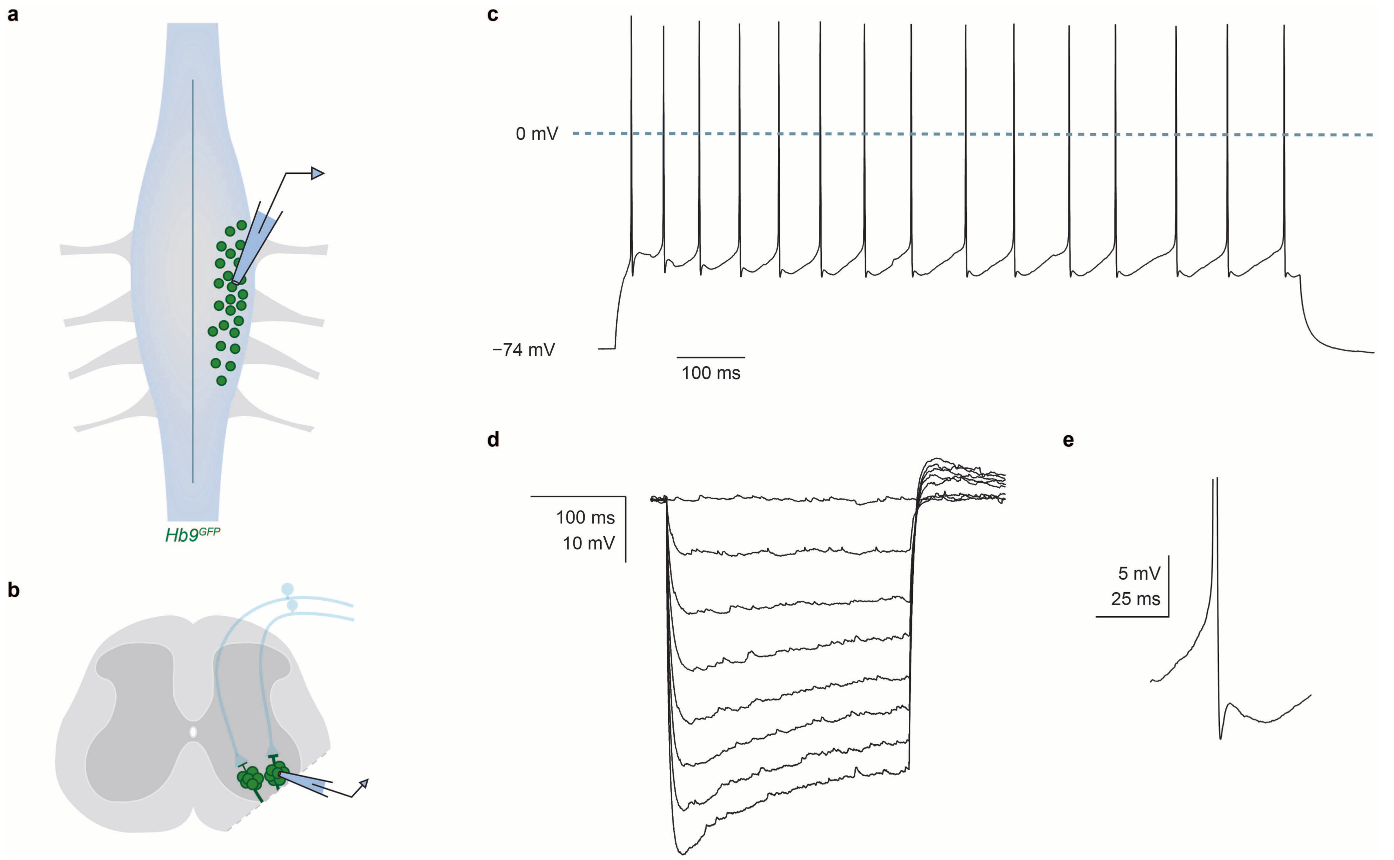
51. Wichterle, H., Lieberam, I., Porter, J. A. & Jessell, T. M. Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110,** 385–397 (2002).
52. Shneider, N. A., Mentis, G. Z., Schustak, J. & O'Donovan, M. J. Functionally reduced sensorimotor connections form with normal specificity despite abnormal muscle spindle development: the role of spindle-derived neurotrophin 3. *J. Neurosci.* **29,** 4719–4735 (2009).
53. Hantman, A. W. & Jessell, T. M. Clarke's column neurons as the focus of a corticospinal corollary circuit. *Nature Neurosci.* **13,** 1233–1239 (2010).
54. Varela, J. A. *et al.* A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J. Neurosci.* **17,** 7926–7940 (1997).
55. Clements, J. D. & Silver, R. A. Unveiling synaptic plasticity: a new graphical and analytical approach. *Trends Neurosci.* **23,** 105–113 (2000).
56. Burke, R. E. & Rudomin, P. in *The Handbook of Physiology: Section 1: The Nervous System* Vol. 1 (ed. Kandel, E. R.) **1,** 877–944 (1977).
57. Bawa, P. & Stein, R. B. Frequency response of human soleus muscle. *J. Neurophysiol.* **39,** 788–793 (1976).



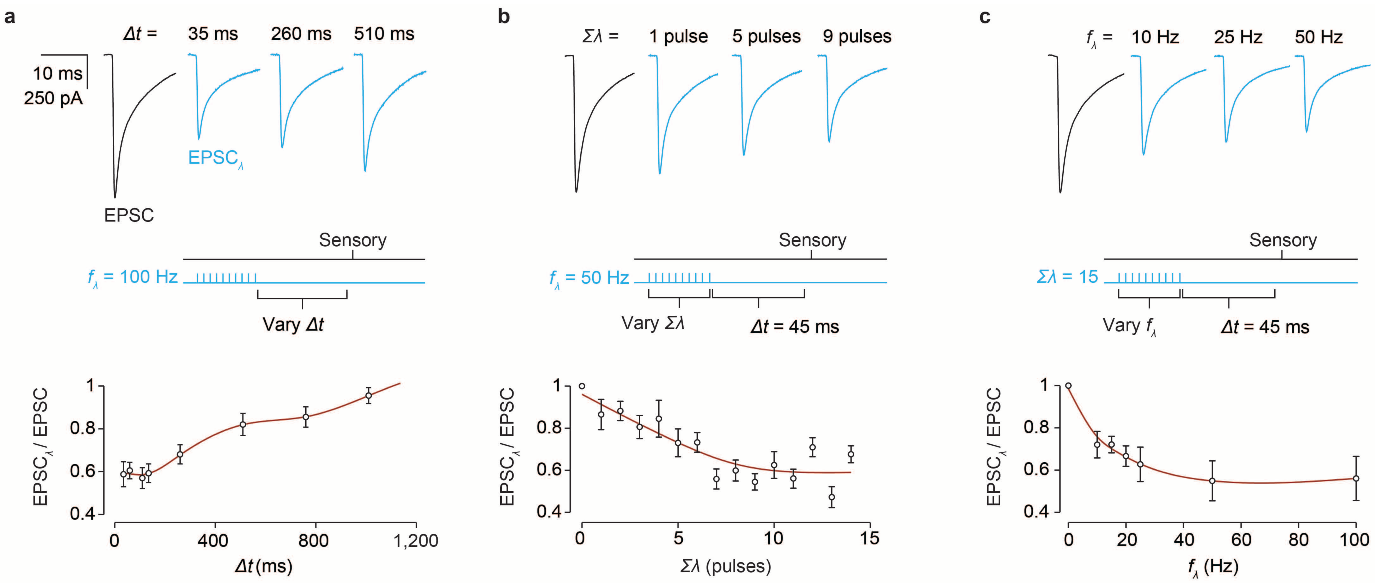
Extended Data Figure 1 | Genetic targeting of GABApre neurons in lumbar spinal cord. **a**, At late ages (p30) *FLEX-ChR2-YFP* lumbar injection in $Gad2^{Cre}$ mice marks GABApre neurons. In ventral horn (yellow box): $YFP^+/GAD2^+$ GABApre boutons contact $vGluT1^+$ proprioceptor terminals (**b**, higher magnification). **c**, $YFP^{OFF}/GAD1^+$ GABApost boutons do not contact $vGluT1^+$ terminals. In contrast, a single $YFP^+/GAD1^+$ bouton is in contact with a $vGluT1^+$ terminal and is, therefore, a GABApre bouton. p30 injection marks GABApre boutons (75.8 ± 3.3%) but not GABApost boutons (0.99 ± 0.02%; $n = 2$). Values indicate mean ± s.e.m.



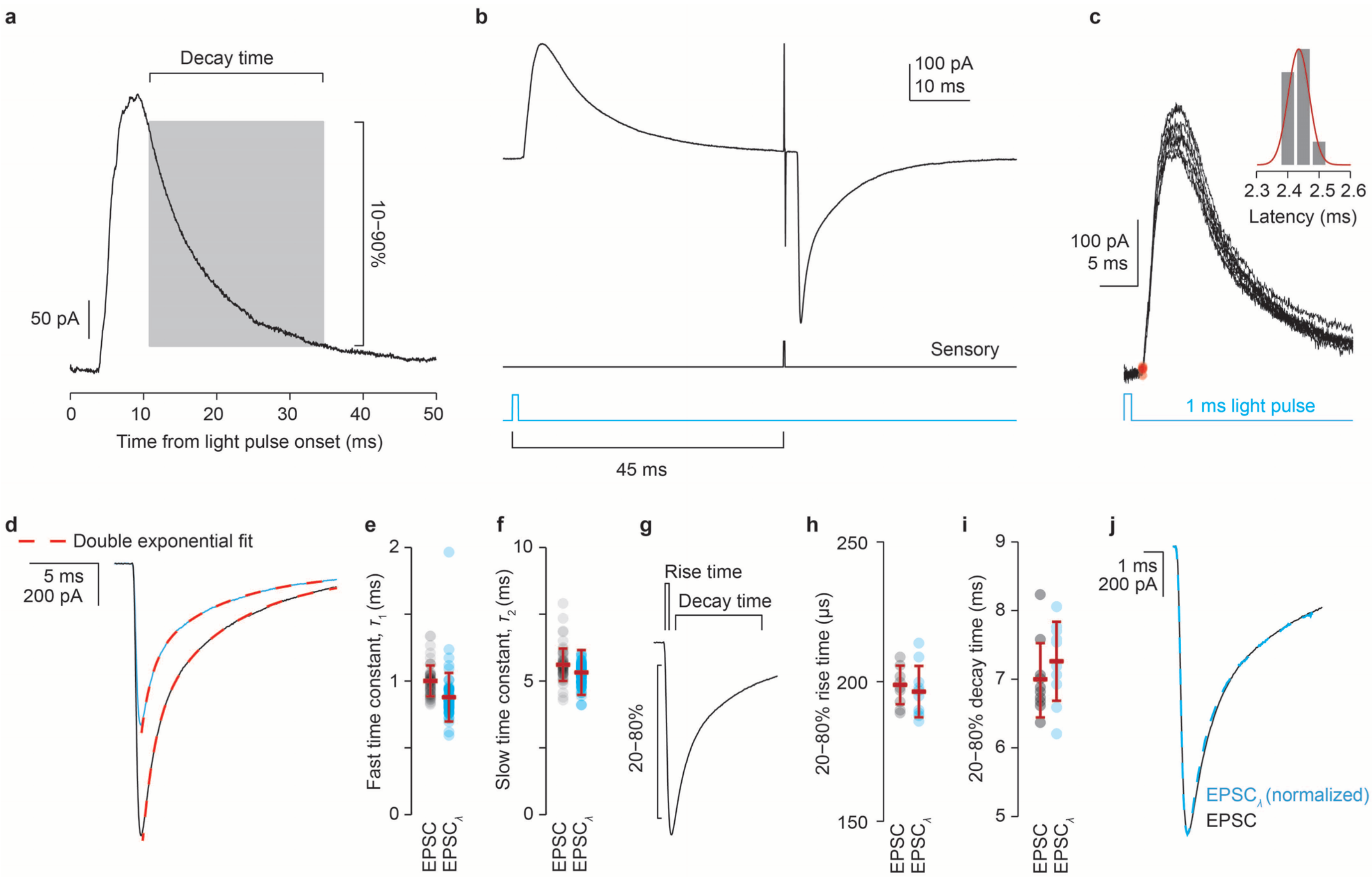
Extended Data Figure 2 | Light-evoked excitation of spinal motor neurons reflects *Gad2*^{Cre} neuron-mediated proprioceptor depolarization. **a**, Recordings from dorsal and ventral roots during *Gad2*^{Cre} neuron photostimulation at 24–26 °C. **b**, Light pulses induced primary afferent depolarization (PAD, top), accompanied by antidromic action potentials (black arrow), in dorsal roots shortly followed by orthodromic discharge in ventral roots (bottom). Outline indicates region shown in **c**. As shown in Fig. 2f, the time course and amplitude of PAD evoked by sensory stimulation and *Gad2*^{Cre}-neuron photostimulation are similar. Sensory-PAD: amplitude 203 ± 2 μV; 10–90% rise time 18.8 ± 0.3 ms; 10–90% decay time 275.7 ± 8.3 ms; $n = 10$ trials; *Gad2*^{Cre}-evoked PAD: amplitude 230 ± 1 μV; 10–90% rise time 17.3 ± 0.3 ms; 10–90% decay time 186.6 ± 2.0 ms. **c**, Mean antidromic spike recorded in dorsal root (top) and orthodromic spike recorded in ventral root (bottom). The latencies of spike onset (outlined region) are shown on the right (dorsal root spike onsets, blue circles; ventral root spike onsets, brown circles; vertical lines indicate latencies from light pulse onset). Ventral root spike onset was consistently later than dorsal root spike onset (mean delay 0.73 ± 0.03 ms; mean dorsal root spike latency 8.01 ± 0.42 ms; mean ventral root spike latency 8.74 ± 0.40 ms; two-tailed paired t test, $P < 0.05$, $n = 2$ preparations), indicating that light-evoked spikes occur in sensory neurons before motor neurons. **d**, Whole-cell patch-clamp recording from spinal motor neurons during *Gad2*^{Cre} neuron photoactivation. **e**, Light-induced excitatory postsynaptic potentials (EPSPs) (1 ms pulses) evoked at 24 (left) but not 30 °C (right). Inset, left, temperature dependence of light-evoked EPSPs (top plot, bath temperature (T); bottom plot, EPSP amplitudes; both plotted as functions of time; note EPSP recovery with return to low temperature). Inset, right, average EPSP amplitude as a function of bath temperature. **f**, During low temperature (24–26 °C) recording conditions, application of the GABA-A receptor antagonist gabazine (SR 95531, Gbz, 2 μM) abolishes light-evoked EPSPs but leaves the predominantly glycinergic light-evoked inhibitory postsynaptic potential (IPSP) intact. Inset, individual light-evoked EPSP amplitudes during gabazine application (Gbz, grey bar). **g**, During low temperature (24–26°) recording conditions, application of the AMPA-receptor antagonist NBQX (2 μM) abolishes the fast, high-amplitude component of the light-evoked EPSP but does not affect the light-evoked IPSP. Inset, individual EPSP amplitudes during NBQX application (NBQX, grey bar). Together these experiments are consistent with the view that synchronous activation of *Gad2*^{Cre} neurons at low temperature depolarizes sensory afferent terminals with sufficient strength to generate sensory action potentials and subsequent glutamate release, in turn activating motor neurons (Supplementary Note 2). Values and error bars indicate mean ± s.e.m.



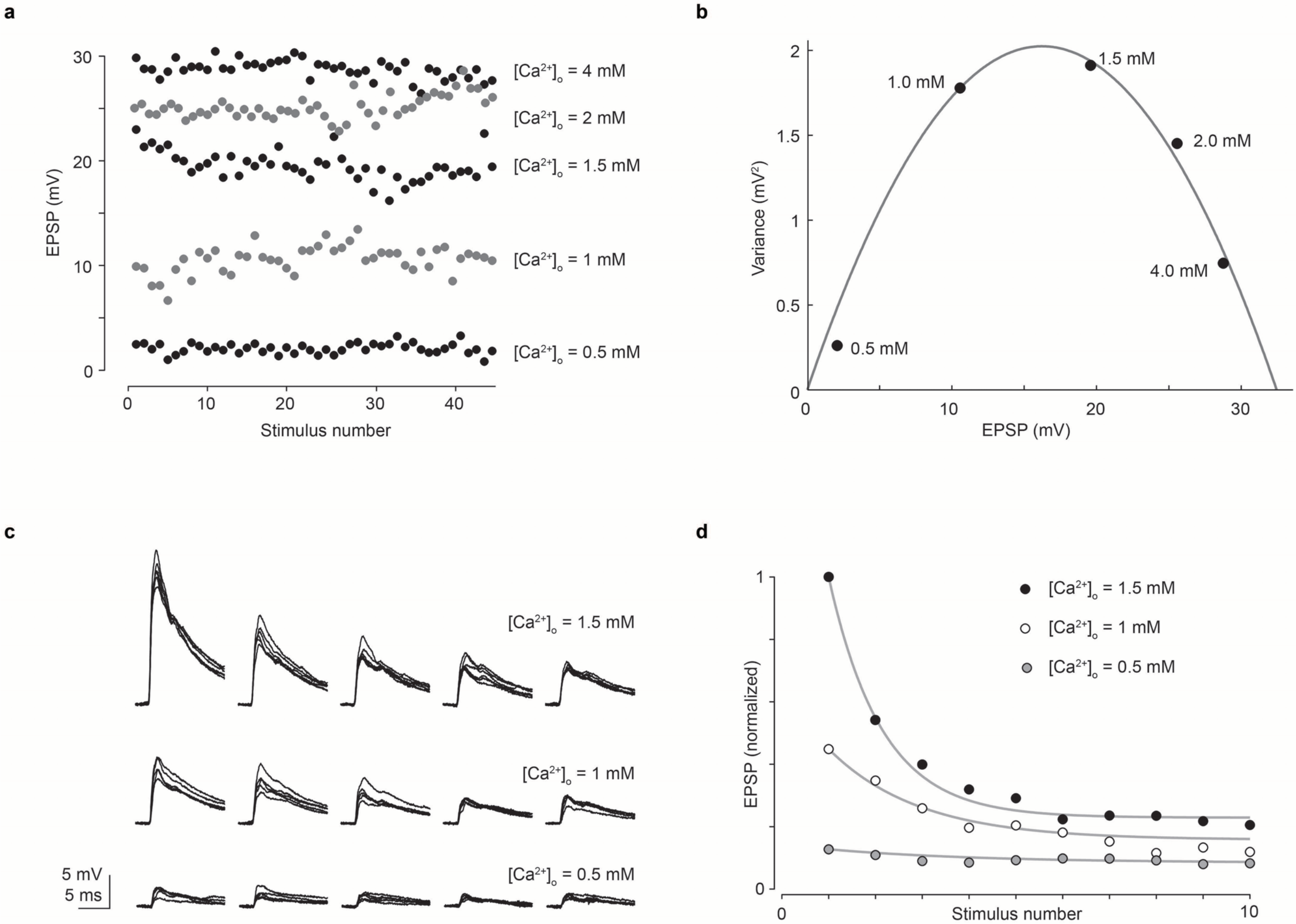
Extended Data Figure 3 | Physiological identification of spinal motor neurons. **a**, Whole-cell patch-clamp recording from spinal motor neurons in a whole (or hemisected) cord, in which motor neurons were targeted via visual guidance using GFP expression in *Hb9*^{GFP} mice. **b**, Removal of the ventral-lateral white matter (see Methods) permits visual identification and access to spinal motor neurons without disrupting sensory–motor or interneuron circuitry. **c–e**, Motor neurons recorded in this configuration exhibit physiological properties typically associated with spinal motor neurons[56]: **c**, current injection elicits repetitive action potentials, with rate adaptation; **d**, hyperpolarizing current steps reveal membrane potential sag, indicative of I_h current; and **e**, the waveform of motor neuron action potentials exhibits an early hyperpolarization indicative of an I_A current and a prolonged hyperpolarization likely mediated by a calcium-activated $I_{K(Ca)}$ current.



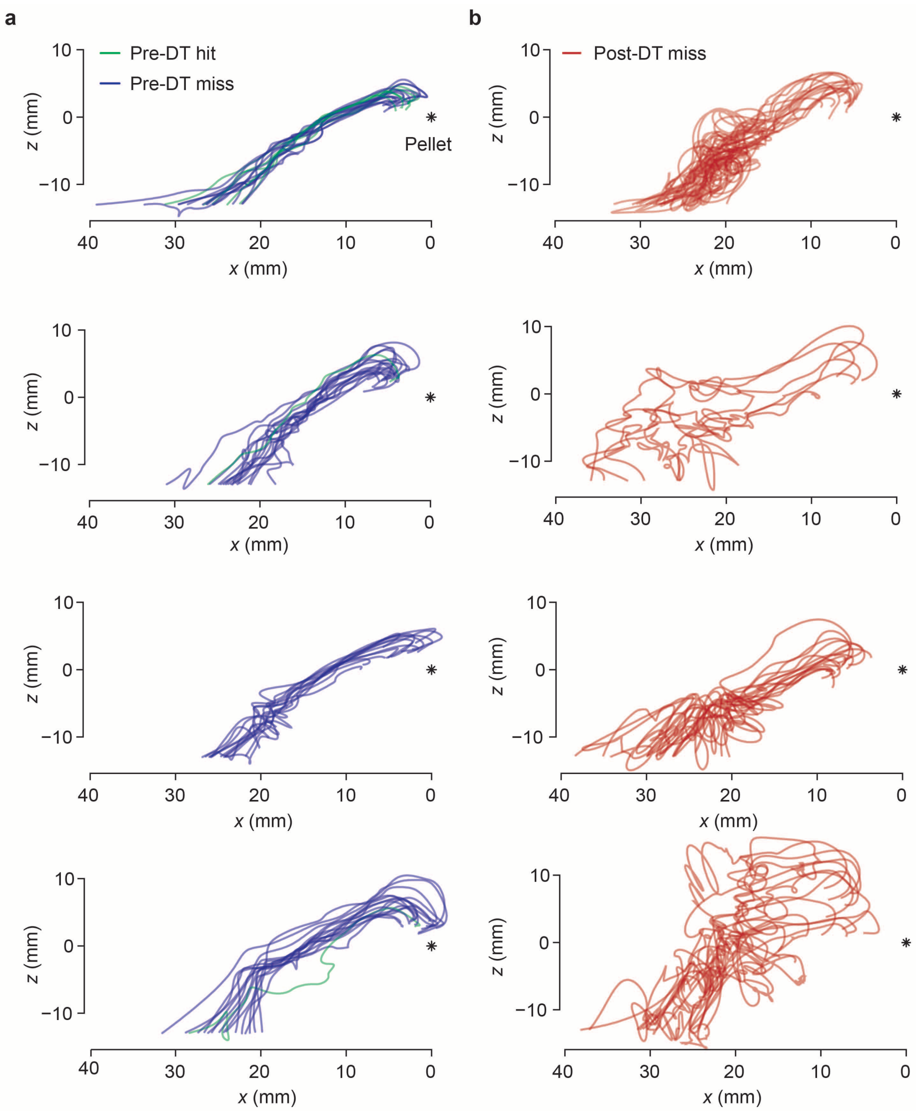
Extended Data Figure 4 | *Gad2^{Cre}* neuron photoactivation scales sensory EPSC amplitude. **a**, Sensory-evoked EPSCs observed in motor neurons at varying delays (Δt) between the cessation of photostimulation (ten 1 ms pulses, 100 Hz, blue line) and sensory stimulation (black line). Fractional EPSC amplitude ($EPSC_\lambda$/EPSC) plotted as a function of sensory delay Δt. Maximal EPSC suppression was seen at the smallest Δt values. Plot depicts fractional EPSC amplitude for Δt = 35, 60, 110, 135, 260, 510, 760 and 1,010 ms. **b**, Sensory-evoked EPSCs following a varying number of light pulses ($\Sigma\lambda$, blue line) at a fixed latency (Δt = 45 ms) recruits *Gad2^{Cre}* neuron-evoked EPSC suppression of increasing magnitude. Plot depicts fractional EPSC amplitude for $\Sigma\lambda$ = 1–14 pulses. **c**, Sensory-evoked EPSCs following varying *Gad2^{Cre}* neuron photoactivation frequency (f_λ, blue line) at a fixed sensory delay (Δt = 45 ms) recruits increasing suppression of sensory-EPSCs. Plot depicts fractional EPSC amplitude for f_λ = 10, 15, 20, 25, 50 and 100 Hz. These data, along with the linear scaling of *Gad2^{Cre}* neuron firing with photoactivation frequency (Fig. 2d), indicate that the graded suppression of sensory-EPSCs is a consequence of increased spiking in *Gad2^{Cre}* neurons. Red curves indicate splines fit to the data. Error bars indicate mean ± s.d. We note that these experiments require effective control of *Gad2^{Cre}* neuronal spiking. As shown in Fig. 2b–d, light pulses induced pronounced currents and reliable spiking in ChR2–YFP$^+$ *Gad2^{Cre}* neurons ($V_{holding}$ −60 mV; peak 773 ± 268 pA; steady state 537 ± 193 pA; 10–90% rise time 2.1 ± 0.4 ms; 10–90% decay time 33.5 ± 4.3 ms; n = 3)[20]. Values indicate mean ± s.e.m.



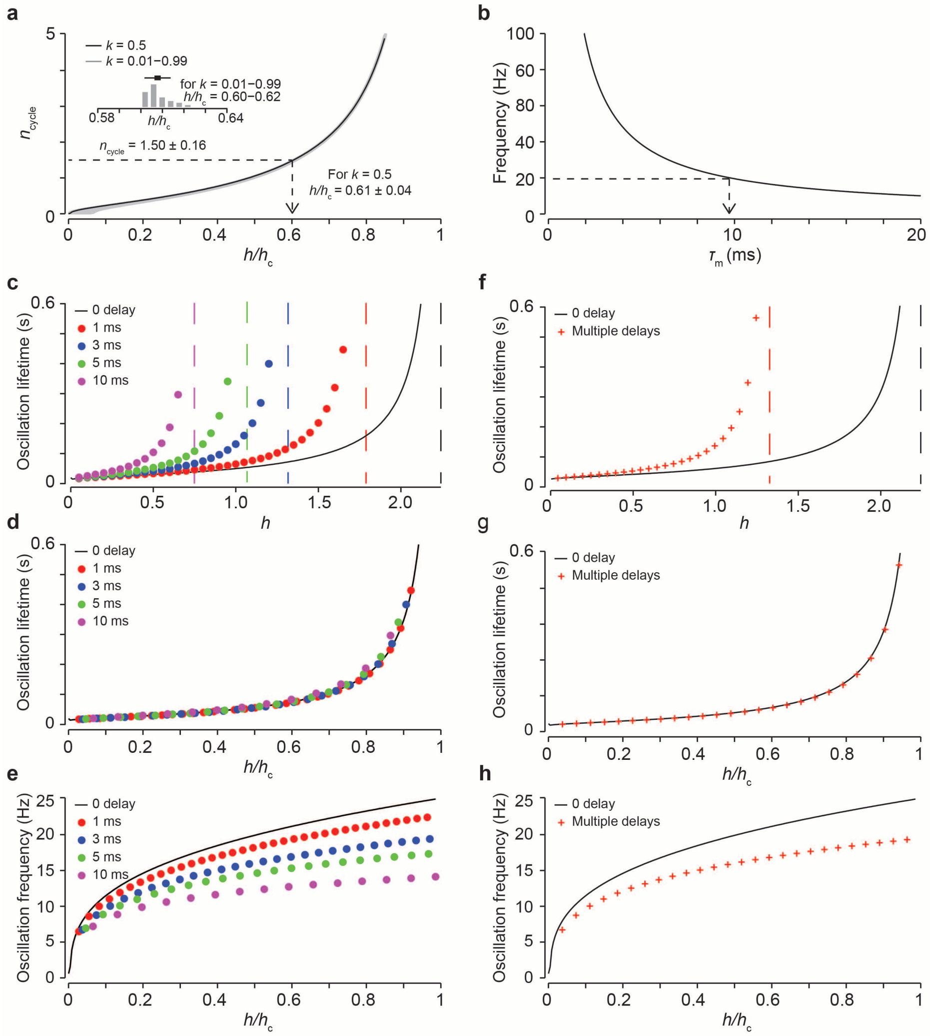
Extended Data Figure 5 | Postsynaptic inhibition cannot account for the observed reduction in sensory-evoked EPSC amplitude. **a**, Motor neuron IPSC (mean of ten trials) elicited by a 1 ms light pulse. IPSCs had a 10–90% decay time of 23.7 ± 1.9 ms and a mean peak amplitude of 299 ± 20 pA at −60 mV holding potential ($n = 9$). IPSCs reversed at ∼−65 mV (Nernst equation predicts $E_{Cl} = -67.5$ mV). **b**, A 45 ms delay between the light pulse (blue) and electrical sensory stimulation pulse (black) permits the GABApost-mediated IPSC to decay almost completely before sensory-EPSC onset. **c**, *Gad2^Cre* neuron-evoked IPSCs are monosynaptic. Red circles indicate estimated onset times with distribution of individual IPSC latencies shown in inset. Mean IPSC onset latency 2.4 ± 0.3 ms; cv_{onset} 0.03 ± 0.007; $n = 5$. As shown in Fig. 3b–d, IPSCs are almost completely blocked by strychnine (Str), with only limited effects of GABA receptor antagonists (Gbz/Cgp). IPSC index, ratio of IPSC amplitude with drug to control IPSC (Str 0.09 ± 0.02, two-tailed paired *t* test, $P < 10^{-5}$, $n = 5$; Gbz/Cgp 0.82 ± 0.04, $P < 0.05$, $n = 3$). Moreover, as shown in Fig. 3e–g, the *Gad2^Cre* neuron-evoked suppression of sensory-evoked EPSCs is unaffected by strychnine (Str) but abolished by co-application of the GABA-A and -B receptor antagonists SR 95531 and CGP 54626. EPSC index, drug has no effect = 1; drug abolishes EPSC-suppression = 0 (Str 1.04 ± 0.02, two-tailed paired *t* test, $P = 0.3$, $n = 3$; Gbz/Cgp 0.02 ± 0.03, two-tailed paired *t* test, $P < 0.02$, $n = 4$; see Methods). **d**, Sensory-evoked EPSC (black trace, control) and $EPSC_\lambda$ (blue trace, with *Gad2^Cre* photoactivation) waveforms fit with double exponential functions ($EPSC_{fit}$, dashed lines, red) of the form $EPSC_{fit}(t) = A1\exp(-t/\tau_1) + A2\exp(-t/\tau_2)$. **e**, Distribution of the fast decay time constant (τ_1) is similar for EPSC (black circles) and $EPSC_\lambda$ (blue circles) waveforms (EPSC τ_1 96 ± 13 μs; $EPSC_\lambda$ τ_1 89 ± 14 μs; $n = 100$ trials). **f**, Distribution of the slow decay time constant (τ_2) is similar for EPSC and $EPSC_\lambda$ waveforms (EPSC τ_2 5.56 ± 0.41 ms; $EPSC_\lambda$ τ_2 5.41 ± 0.48 ms). **g**, Estimation of the 20–80% rise and decay time provides an independent measure of EPSC kinetics. **h**, The 20–80% rise time for control EPSCs (EPSC 199 ± 7 μs) and that of EPSCs evoked following photoactivation ($EPSC_\lambda$ 197 ± 9 μs). **i**, The 20–80% decay time for control EPSCs (EPSC 7.0 ± 0.5 ms) and that of EPSCs evoked following photoactivation ($EPSC_\lambda$ 7.3 ± 0.6 ms). **j**, Normalization of the light-conditioned $EPSC_\lambda$ waveform to the unconditioned EPSC waveform permits estimation of waveform correlation as a further metric of the similarity of the two waveforms (Pearson's correlation, $R > 0.999$, $P < 0.0001$). The unaltered time course of the EPSC waveform argues against light-evoked EPSC suppression mediated by a postsynaptic inhibitory conductance. Although in theory a postsynaptic inhibitory conductance should have no effect on the current waveform of a perfectly clamped EPSC, the large size and low input resistance of motor neurons, as well as the dendritic nature of the majority of sensory inputs, makes it likely that the voltage clamp is insufficient to prevent some depolarization of the dendrites, which would be altered via inhibitory conductance. In support of this idea the time course of EPSP waveforms recorded in current clamp are unaltered by photoactivation of *Gad2^Cre* neurons under the conditions described here (data not shown). Values and error bars indicate mean ± s.e.m. in **a**–**c** and mean ± s.d for **d**–**j**.



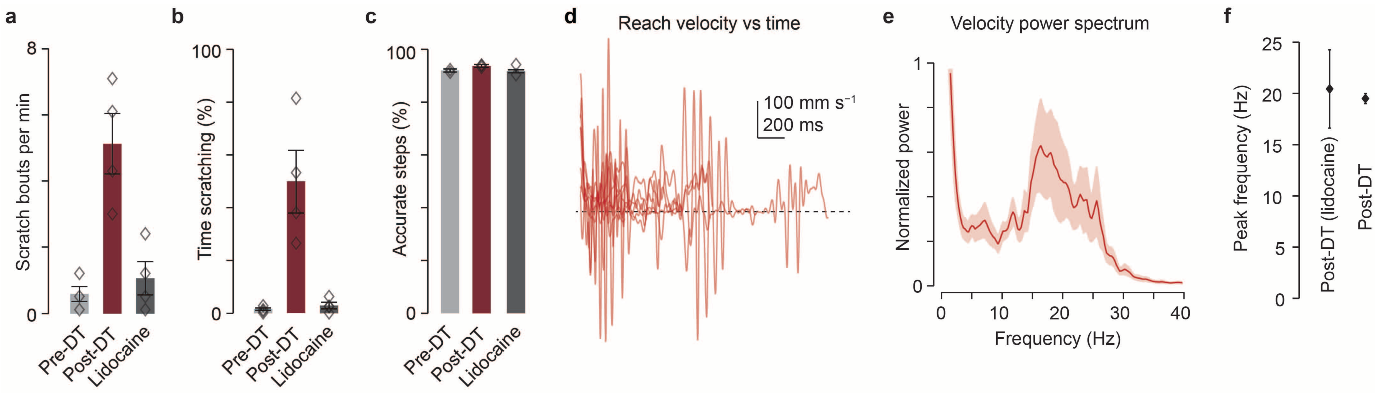
Extended Data Figure 6 | Estimation of release probability using mean variance analysis validates the short-term depression model. **a**, Amplitudes of 225 EPSPs recorded from a spinal motor neuron evoked by sensory stimulation across five calcium concentrations (0.5, 1.0, 1.5, 2.0 and 4.0 mM) exhibiting fluctuating amplitudes due to calcium-controlled changes in transmitter release probability. EPSPs were recorded at each calcium concentration (0.1 Hz sensory stimulation frequency) once EPSP amplitude had settled to a steady value. Each calcium concentration was visited twice in sequence per recording. Data were collected in current clamp to reduce possible confounds due to changes in series resistance over the long recordings required to collect the data. **b**, EPSP variance versus amplitude (black circles) is well fit by a quadratic function[55] (grey line) of the form $\sigma^2 = A\mu + B\mu^2$, where σ^2 is the variance across EPSPs and μ is the mean EPSP amplitude for each calcium condition. The release probability (p_r) is given by $p_r = \mu\,(B/A)\,(1 + cv_q^2)$, where cv_q is the coefficient of variation in EPSP amplitude from a single release site. We set $cv_q = 0.3$, following accepted values[55]. Using this method, we found that for the representative neuron shown here $p_r = 0.97$, 0.86, 0.66, 0.35 and 0.07 for $[Ca^{2+}]_o = 4$, 2, 1.5, 1 and 0.5 mM, respectively. **c**, At each calcium concentration EPSP depression was monitored via repetitive stimulation of the dorsal root (10 pulses, 10 Hz). Shown here are five consecutive EPSPs for three calcium concentrations (1.5, 1.0 and 0.5 mM; superposition of six trials). **d**, Changes in EPSP depression as a function of calcium concentration were quantified using a model of short-term depression[29,54] (see Fig. 3h–j). The short-term depression model was set to generate EPSP amplitudes that depressed at different rates because of changes to the release probability term of the model. These EPSP amplitudes were fit by exponential functions. Normalized mean EPSP amplitudes at different calcium concentrations (circles) were then superimposed upon the curves generated by the short-term depression model (grey lines). The depression shown here is captured by setting the release probability parameters in the short-term depression model to $p_r = 0.52$, $p_r = 0.30$ and $p_r = 0.07$ for $[Ca^{2+}]_o = 1.5$, 1.0 and 0.5 mM, respectively, which are in good correspondence with the release probability terms estimated using mean variance analysis. For $[Ca^{2+}]_o > 1.5$ mM the short-term depression model consistently produced lower estimates for release probability than the mean variance analysis. But the correspondence at 1.5 mM and below argues for the short-term depression model being an effective means of capturing changes in release probability at sensory–motor synapses.



Extended Data Figure 7 | Reach trajectories before and after ablation of GABApre neurons. **a**, Pre-DT reach trajectories for four mice. Two dimensional paw trajectories were generated by projecting the three dimensional trajectory into the *x*–*z* plane (see Fig. 5a). Green traces represent pre-DT hits, blue traces, pre-DT misses. The asterisk indicates the location of the target food pellet. **b**, Post-DT two dimensional reach trajectories for the same mice, red traces. These mice were used for all kinematic quantification included in Extended Data Table 1.



Extended Data Figure 8 | Fitting the model to the frequency and decay time of post-DT reach velocity oscillations. **a**, To fit the model parameters to the experimentally-derived frequency and decay times, we took advantage of the fact that the product of the oscillatory frequency and decay time ($f_{osc}\tau_{osc}$) creates a parameter free value n_{cycle}, which does not depend on the muscle time constant. The term n_{cycle} corresponds to the number of oscillatory cycles per decay time constant and scales as a function of feedback gain. Plotting n_{cycle} as a function of feedback gain (h/h_c, where h_c is the critical gain above which oscillations do not decay) permits estimation of the gain value that corresponds to the experimentally-derived value of n_{cycle}. Importantly, the corresponding gain value does not depend on the drag parameter k (inset and grey shading beneath black curve). Varying k by 100-fold results in minimal changes in the corresponding gain, with h/h_c varying from 0.60 to 0.62. The grey curves represent the relationship between n_{cycle} and h/h_c across the full range of k values. Arrow indicates the gain value corresponding to the experimentally-derived value of n_{cycle} for $k = 0.5$. See Supplementary Note 4 and Methods. **b**, To estimate the appropriate muscle time constant (τ_m), we used the five dimensional matrix that defines the model (see Methods) to calculate the imaginary part of the complex eigenvalue corresponding to $k = 0.5$ and $h/h_c = 0.61$. We found a time constant of $\tau_m = 9.6$ ms (indicated by the arrow), which is within the range used in the original model[11] based on experimentally-derived values[57]. **c**, Introduction of delays in the feedback loop does not alter the basic properties of the model. Oscillation lifetime (τ_{osc}) as a function of absolute gain h for delays of 1, 3, 5 and 10 ms as compared to no delay (black trace). The vertical dashed lines indicate the gain level above which oscillations never decay (the critical gain, h_c) for each delay value. **d**, Scaling these curves by their corresponding critical gain values (h/h_c) reveals equivalent oscillatory lifetimes. **e**, Oscillation frequency (f_{osc}) as a function of normalized gain. As feedback delays increase, peak oscillatory frequency decreases. For higher delay conditions, where the oscillation frequency is significantly below 20 Hz, decreasing the muscle time constant results in higher oscillation frequencies. Thus for longer delays ~20 Hz oscillation is possible, but requires an increasingly small muscle time constant. **f–h** are as **c–e** but with three simultaneous loops of different delays (1, 3 and 5 ms). In the presence of multiple simultaneous feedback delays the model continues to oscillate at a single dominant peak frequency for a given gain level. See Supplementary Note 5 and Methods.



Extended Data Figure 9 | Lidocaine application abolishes scratching behaviour but not reach oscillations, and does not affect forepaw stepping behaviour. **a**, Scratching behaviour increased following GABApre neuronal ablation, but was reduced to normal levels by topical application of lidocaine to the right forearm and paw. The rate of scratching bouts per 10 min observation session increased following DT-administration, but returned to normal levels following lidocaine application (mean rate of scratch bouts per minute: pre-DT $0.6 \pm 0.2\,\mathrm{min}^{-1}$; post-DT $5.1 \pm 0.9\,\mathrm{min}^{-1}$; post-DT plus lidocaine $1.0 \pm 0.5\,\mathrm{min}^{-1}$; two-tailed paired *t*-test, pre-DT versus post-DT, $P < 0.05$; post-DT versus post-DT + lidocaine, $P < 0.05$, $n = 4$). **b**, Similarly, the percent time spent scratching during a 10 min observation period is normally low (pre-DT 1.6 ± 0.4%) but increased following DT-administration (post-DT 49.8 ± 11.7%) and trended towards baseline after lidocaine application (post-DT plus lidocaine 2.9 ± 1.3%). Two-tailed paired *t*-test, pre-DT versus post-DT, $P < 0.05$; post-DT versus post-DT + lidocaine, $P = 0.06$, $n = 4$. **c**, Stepping success rate of the right forepaw on the horizontal ladder task[33]. Lidocaine did not affect stepping performance (pre-DT 90 ± 1%; post-DT 93 ± 1%; post-DT + lidocaine 92 ± 1%; two-tailed paired *t* test, pre-DT versus post-DT, $P = 0.7$; post-DT versus post-DT + lidocaine, $P = 0.6$, $n = 3$). The equivalent rate of accuracy across conditions indicates that lidocaine application has no overt effect on forepaw placement during stepping. As shown in Fig. 4f, right and left forepaw placement accuracy pre-DT versus post-DT were similar (right paw: pre-DT 90.2 ± 2.0%; post-DT 79.6 ± 13.2%; two-tailed paired *t* test, $P = 0.41$; left paw: pre-DT 80.8 ± 2.7%; post-DT 76.1 ± 6.9%; two-tailed paired *t* test, $P = 0.34$, $n = 4$.) In contrast to stepping accuracy, as shown in Fig. 4e, reach success degraded following GABApre ablation as compared to control mice[31] (pre-DT 48.6 ± 3.7%; post-DT 4.9 ± 4.7%; two-way repeated-measures ANOVA, interaction of group × toxin: $F_{1,6} = 17.64$, $P = 0.006$; post hoc Bonferroni test, DTR: $P < 0.01$, $n = 4$ DTR, $n = 4$ control). **d**, Post-DT velocities of individual reaches from a representative mouse continued to exhibit oscillation following topical lidocaine application. **e**, Power spectrum of post-DT reaches following lidocaine application ($n = 2$ mice, 14 reaches; shaded area, s.d.). **f**, Mean dominant frequency peak for post-DT reaches with lidocaine (20.5 ± 3.8 Hz) and without lidocaine (19.5 ± 0.5 Hz; see Fig. 5f,g). The persistence of limb oscillation during lidocaine block implicates a loss of proprioceptive rather than cutaneous presynaptic inhibition as the origin of the oscillation uncovered by GABApre neuronal ablation. Values and error bars indicate mean ± s.e.m.

Extended Data Table 1 | Reaching kinematics after GABApre neuronal ablation

	Reach phase		Grab phase	
	Pre-DT	Post-DT	Pre-DT	Post-DT
Direction reversals (no.)	1.38 ± 0.30	13.36 ± 1.36	1.72 ± 0.12	1.18 ± 0.08
Direction reversals at >100 mm s^{-1} (no.)	0.07 ± 0.02	5.68 ± 0.46	0.02 ± 0.01	0.10 ± 0.03
s.d. of distance to target (mm)	2.93 ± 0.24	4.89 ± 0.25	0.91 ± 0.09	0.92 ± 0.05
s.d. of velocity (mm s^{-1})	36.34 ± 1.80	92.05 ± 6.80	35.20 ± 2.10	48.11 ± 2.24
Total distance travelled per reach (mm)	26.87 ± 0.44	84.67 ± 3.14	12.63 ± 0.40	8.03 ± 0.31
Total duration per reach (ms)	255.20 ± 7.84	582.93 ± 47.24	116.47 ± 6.80	65.78 ± 3.96
Mean reach speed (mm s^{-1})	113.79 ± 5.80	157.99 ± 12.03	106.83 ± 4.27	133.25 ± 5.64

	Entire reach	
	Pre-DT	Post-DT
Initial peak velocity at <50 ms (mm s^{-1})	296.96 ± 16.73	284.14 ± 50.61
Subsequent peak velocity at >50 ms (mm s^{-1})	161.80 ± 10.36	274.17 ± 29.06
Peak power (18−22 Hz)	143 ± 37	957 ± 270

We observed an increase in the following kinematic parameters in the reach, but not the grab phase in GABApre ablated relative to control mice[31]: mean number of forepaw direction reversals towards and away from the pellet ($n = 4$ DTR mice, 112 reaches, $n = 4$ control mice, 75 reaches; two-way repeated-measures ANOVA, interaction of group × condition, reach phase: $F_{1,6} = 15.48$, $P = 0.01$; post hoc Bonferroni test, DTR pre-DT versus post-DT, $P < 0.001$; grab phase: $F_{1,6} = 3.24$, $P = 0.12$); mean number of high velocity (>100 mm s^{-1}) reversals of paw direction (reach phase: $F_{1,6} = 35.04$, $P = 0.001$; post hoc Bonferroni test, DTR pre-DT versus post-DT, $P < 0.001$; grab phase: $F_{1,6} = 2.71$, $P = 0.15$); mean s.d. of paw distance to target (reach phase: $F_{1,6} = 8.57$, $P = 0.03$; grab phase: $F_{1,6} = 0.60$, $P = 0.47$); mean s.d. of paw velocity (reach phase: $F_{1,6} = 25.55$, $P = 0.002$; post hoc Bonferroni test, DTR pre-DT versus post-DT, $P < 0.01$; grab phase: $F_{1,6} = 1.12$, $P = 0.33$); the mean total distance travelled by the paw (reach phase increase in distance travelled: $F_{1,6} = 62.42$, $P = 0.0002$; post hoc Bonferroni test, DTR pre-DT versus post-DT, $P < 0.0001$; grab phase decrease in distance travelled: $F_{1,6} = 15.57$, $P = 0.01$; post hoc Bonferroni test, DTR pre-DT versus post-DT, $P < 0.01$); and mean movement duration (reach phase increase in duration: $F_{1,6} = 12.67$, $P = 0.01$; post hoc Bonferroni test, DTR pre-DT versus post-DT, $P < 0.01$; grab phase decrease in duration: $F_{1,6} = 13.14$, $P = 0.01$; post hoc Bonferroni test, DTR pre-DT versus post-DT, $P < 0.01$). Mean reach speed did not differ significantly between conditions. Digit abduction (maximum distance between digits 2 and 4) during grasp attempts was unaffected by GABApre neuronal ablation (pre-DT, 3.38 ± 0.25 mm; post-DT, 3.39 ± 0.24 mm; two-tailed paired t test, $P = 0.98$, $n = 5$). Reaching kinematics were not affected during any phase of movement in DT-treated control mice that did not receive *FLEX-DTR-GFP*[31]. As shown in Fig. 5d,e the amplitude of the first velocity peak (within the first 50 ms of the reach) did not differ between pre-DT and post-DT reaches, while the subsequent peak velocity (>50 ms from reach initiation) was significantly higher for post-DT reaches (initial peak: pre-DT versus post-DT, two-tailed paired t test, $P = 0.82$; subsequent peak: pre-DT versus post-DT, two-tailed paired t test, $P = 0.04$, $n = 4$ mice, 112 reaches). As shown in Fig. 5g, peak power in the 18–22 Hz frequency band was significantly higher in post-DT mice (two-tailed paired t test, $P < 0.02$, $n = 5$ mice, 132 reaches). Successful and unsuccessful pre-DT reaches were grouped together for all analyses. There were no post-DT successful reaches in the kinematic assay. Values indicate mean ± s.e.m.

doi:10.1038/nature13035

Haematopoietic stem cells require a highly regulated protein synthesis rate

Robert A. J. Signer[1], Jeffrey A. Magee[1], Adrian Salic[2] & Sean J. Morrison[1]

Many aspects of cellular physiology remain unstudied in somatic stem cells, for example, there are almost no data on protein synthesis in any somatic stem cell. Here we set out to compare protein synthesis in haematopoietic stem cells (HSCs) and restricted haematopoietic progenitors. We found that the amount of protein synthesized per hour in HSCs *in vivo* was lower than in most other haematopoietic cells, even if we controlled for differences in cell cycle status or forced HSCs to undergo self-renewing divisions. Reduced ribosome function in *Rpl24*$^{Bst/+}$ mice further reduced protein synthesis in HSCs and impaired HSC function. *Pten* deletion increased protein synthesis in HSCs but also reduced HSC function. *Rpl24*$^{Bst/+}$ cell-autonomously rescued the effects of *Pten* deletion in HSCs; blocking the increase in protein synthesis, restoring HSC function, and delaying leukaemogenesis. *Pten* deficiency thus depletes HSCs and promotes leukaemia partly by increasing protein synthesis. Either increased or decreased protein synthesis impairs HSC function.

Mutations in ribosomes and other gene products that affect protein synthesis are associated with human diseases marked by haematopoietic dysfunction[1,2]. Increased protein synthesis can promote the development and progression of certain cancers, including haematopoietic malignancies[3–6]. Ribosomal defects commonly impair HSC and erythroid progenitor function[7–11]. However, it is not clear whether these defects reflect a catastrophic reduction in protein synthesis below the level required for cellular homeostasis or whether HSCs require highly regulated protein synthesis.

Methods for measuring protein synthesis have depended upon the incorporation of radiolabelled amino acids, amino acid analogues[12], or puromycin[13–15] into nascent polypeptides in cultured cells. However, somatic stem cells profoundly change their properties in culture[16] necessitating the analysis of protein synthesis in rare cells *in vivo*. A new fluorogenic assay using O-propargyl-puromycin (OP-Puro) has been developed to image protein synthesis *in vivo*[17]. OP-Puro, like puromycin, is taken up by cells *in vivo*, entering ribosome acceptor sites and incorporating into nascent polypeptides[17]. An azide-alkyne reaction can be used to label OP-Puro fluorescently to quantitate protein synthesis in individual cells[17]. We adapted this approach to quantify protein synthesis by haematopoietic cells using flow cytometry.

HSCs synthesize less protein per hour

We administered a single intraperitoneal injection of OP-Puro (50 mg kg^{-1} body mass) then killed mice 1 h later and isolated bone marrow cells. We did not detect toxicity, signs of illness, changes in bone marrow cellularity, or changes in the frequencies of CD150$^+$CD48$^-$Lineage$^-$ Sca-1$^+$c-kit$^+$ (CD150$^+$CD48$^-$LSK) HSCs[18], annexin V$^+$ bone marrow cells, annexin V$^+$ HSCs, or dividing HSCs (Extended Data Fig. 1a–e).

Bone marrow cells from OP-Puro-treated mice showed a clear increase in fluorescence relative to untreated mice (Fig. 1a). The translation inhibitor cycloheximide profoundly blocked OP-Puro incorporation by bone marrow cells in culture (Fig. 1b). Incorporation of the methionine analogues, L-homopropargylglycine (HPG) and L-azidohomoalanine (AHA), into bone marrow cells, common myeloid progenitors (CMPs), granulocyte-macrophage progenitors (GMPs) and Gr-1$^+$ myeloid cells correlated with OP-Puro incorporation in culture (Fig. 1c–f).

HSCs incorporated less OP-Puro than most other bone marrow cells from the same mice (Fig. 1g). This suggested that HSCs synthesize less protein per hour than most other haematopoietic progenitors. CD150$^-$ CD48$^-$LSK multipotent progenitors (MPPs)[19] showed similar OP-Puro incorporation as HSCs (Fig. 1h); however, the mean rate of OP-Puro incorporation was significantly higher in unfractionated bone marrow cells, CMPs, GMPs, megakaryocyte-erythroid progenitors (MEPs), Gr-1$^+$ myeloid cells, B220$^+$IgM$^-$CD43$^+$ pro-B cells, B220$^+$IgM$^-$CD43$^-$ pre-B cells, B220$^+$IgM$^+$ B cells, CD3$^+$ T cells, and CD71$^+$Ter119$^+$ erythroid progenitors (Fig. 1h). Extended Data Figure 1f–i shows markers, gating strategies and OP-Puro incorporation histograms for each cell population.

To test whether reduced OP-Puro incorporation into HSCs reflects OP-Puro efflux by the Abcg2 transporter (also known as Bcrp1) we administered OP-Puro to *Abcg2*-deficient mice, which lack efflux activity in HSCs[20]. *Abcg2*$^{-/-}$ HSCs continued to show significantly lower mean rates of OP-Puro incorporation compared to most other *Abcg2*$^{-/-}$ progenitors (Fig. 2a), similar to the lowest levels observed among bone marrow cells (Fig. 2b).

Differences in OP-Puro incorporation did not reflect proteasomal degradation[21]. The maximum OP-Puro signal in haematopoietic cells was 1 h after OP-Puro administration (data not shown). However, HSCs showed little decline in OP-Puro signal between 1 and 3 h after administration and had significantly less OP-Puro incorporation than any restricted progenitor at both times (Fig. 2c). In contrast, the OP-Puro signal was profoundly reduced 24 h after administration (Fig. 2c). This suggested that degradation of OP-Puro-containing polypeptides requires several hours. Consistent with this, incubation of bone marrow cells at 37 °C for 30 min did not significantly reduce OP-Puro fluorescence in any cell population relative to an aliquot of the same cells kept on ice to arrest degradation (Extended Data Fig. 2a).

We pre-treated mice with bortezomib (1 mg kg^{-1}; intravenous) to inhibit proteasome activity[22] 1 h before OP-Puro administration. This significantly increased the OP-Puro signal in nearly all cell populations 24 h after OP-Puro administration (Extended Data Fig. 2b). When we assessed OP-Puro incorporation 1 h after administration, bortezomib pre-treatment only modestly increased OP-Puro fluorescence in certain

[1]Howard Hughes Medical Institute, Children's Research Institute, Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [2]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

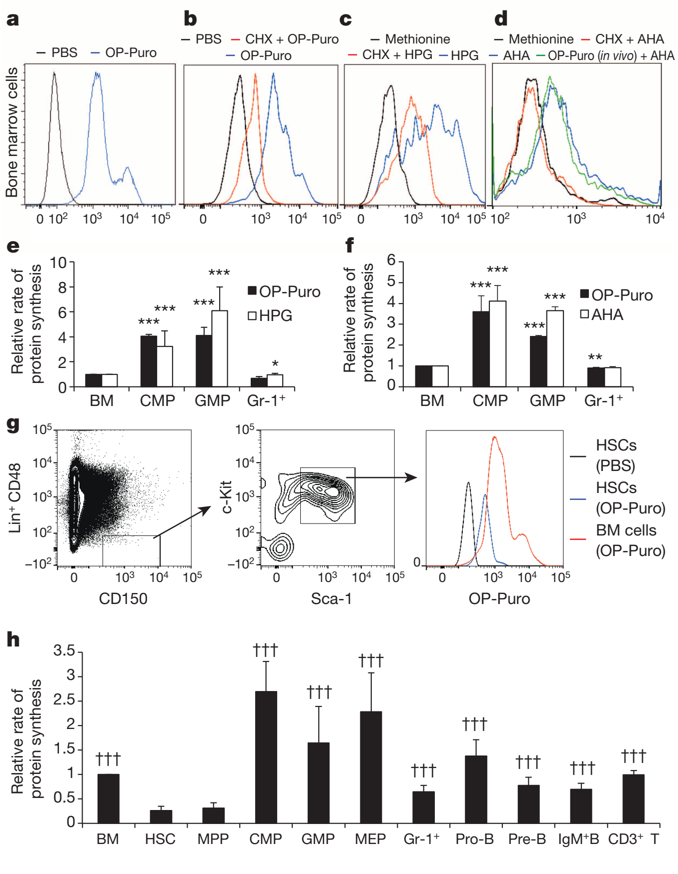


Figure 1 | Quantification of protein synthesis in haematopoietic cells *in vivo*. **a**, OP-Puro incorporation in bone marrow (BM) cells *in vivo* 1 h after administration. **b–d**, OP-Puro (**b**), HPG (**c**), and AHA (**d**) incorporation in bone marrow cells in culture was inhibited by cycloheximide (CHX). **d**, Bone marrow cells from mice treated with OP-Puro *in vivo* exhibited normal AHA incorporation in culture, indicating that OP-Puro did not block protein synthesis. **e**, **f**, OP-Puro versus HPG (**e**; $n = 4$ mice from 2 experiments) or AHA (**f**; $n = 3$ mice from 3 experiments) incorporation by haematopoietic cells in culture. **g**, OP-Puro incorporation in $CD150^+CD48^-$LSK HSCs and unfractionated bone marrow cells 1 h after administration *in vivo*. **h**, Protein synthesis in various haematopoietic stem cell and haematopoietic progenitor cell populations relative to unfractionated bone marrow cells ($n = 15$ mice from 9 experiments). Extended Data Fig. 1j shows the data from Fig. 1h using a $\log_2$ scale. Data represent mean ± s.d. Statistical significance was assessed using a two-tailed Student's *t*-tests (**e**, **f**) and differences relative to HSCs (**h**) were assessed using a repeated-measures one-way analysis of variance (ANOVA) followed by Dunnett's test for multiple comparisons. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ relative to bone marrow; †††$P < 0.001$ relative to HSCs in **h**.

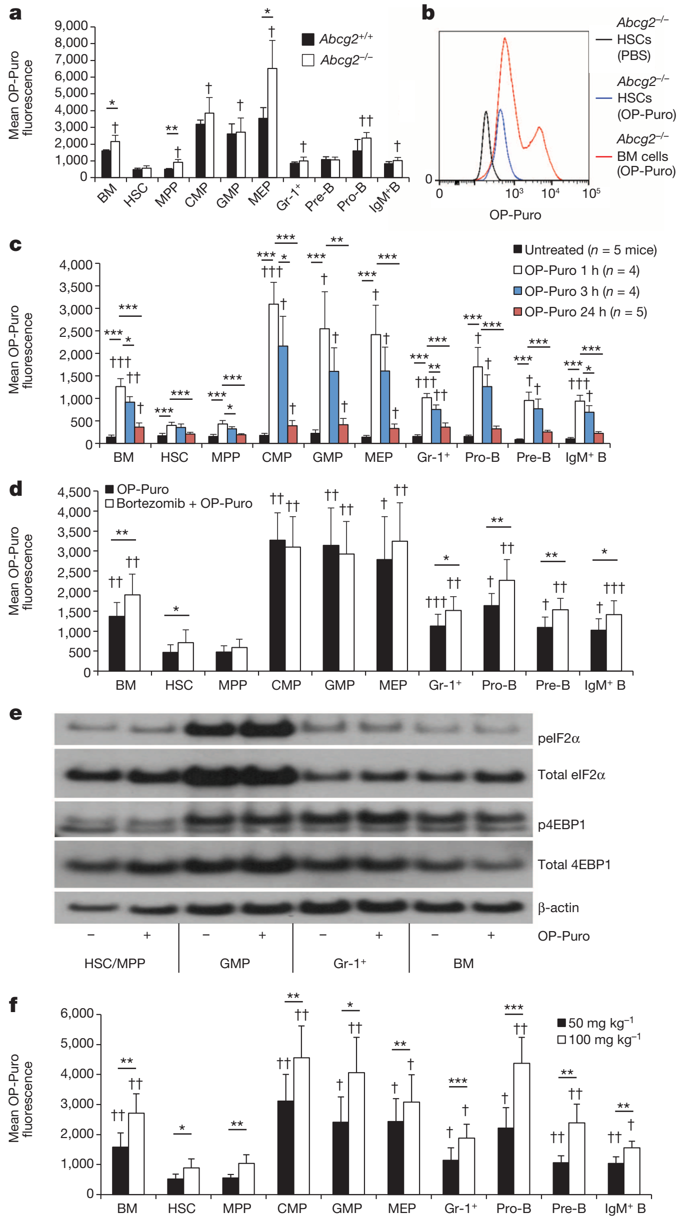


Figure 2 | Lower rate of OP-Puro incorporation by HSCs does not reflect efflux or proteasomal degradation. **a**, **b**, OP-Puro fluorescence in haematopoietic cells from *Abcg2*-deficient and control mice 1 h after OP-Puro administration *in vivo* ($n = 4$ mice from 3 experiments). **c**, OP-Puro fluorescence in haematopoietic cells 1, 3 or 24 h after OP-Puro administration ($n = 5$ experiments). **d**, OP-Puro fluorescence in haematopoietic cells 2 h after bortezomib and 1 h after OP-Puro administration *in vivo* ($n = 5$ mice per treatment from 5 experiments). **e**, Western blot analyses of 30,000 cells from each haematopoietic-cell population from OP-Puro-treated or control mice. **f**, OP-Puro fluorescence in haematopoietic cells 1 h after administering $50\ \mathrm{mg\,kg^{-1}}$ or $100\ \mathrm{mg\,kg^{-1}}$ OP-Puro ($n = 5$ mice per dose from 5 experiments). All data represent mean ± s.d. To assess the statistical significance of treatment effects within the same cells we performed two-tailed Student's *t*-tests; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Differences between HSCs and other cell populations were assessed with a repeated-measures one-way ANOVA followed by Dunnett's test for multiple comparisons; †$P < 0.05$, ††$P < 0.01$, †††$P < 0.001$.

cell populations (Fig. 2d). Even in the presence of bortezomib, HSCs had significantly less OP-Puro fluorescence than restricted progenitors. The lower OP-Puro incorporation by HSCs seems to primarily reflect reduced protein synthesis rather than accelerated proteasomal degradation.

OP-Puro also did not affect the levels of phosphorylated 4EBP1 or eIF2α (Fig. 2e), key regulators of translation that can be influenced by proteotoxic stress[23,24]. HSCs and MPPs had less phosphorylated 4EBP1 and β-actin per cell than other haematopoietic cells but similar levels of total 4EBP1. This raised the possibility that HSCs synthesize less protein as a consequence of increased 4EBP1-mediated inhibition of translation (Fig. 2e).

Doubling the dose of OP-Puro to $100\ \mathrm{mg\,kg^{-1}}$ significantly increased OP-Puro incorporation in all cell populations, but HSCs continued to have significantly lower levels than all other cell populations except MPPs (Fig. 2f). Thus, $50\ \mathrm{mg\,kg^{-1}}$ OP-Puro did not significantly attenuate protein synthesis and uptake was not saturated in HSCs or most other cells.

Dividing HSCs also make less protein

Haematopoietic progenitors in the replicating phases of the cell cycle (S, G_2 and M phases; $S/G_2/M$) exhibited significantly higher rates of OP-Puro incorporation than cells in G_0 and G_1 phases (G_0/G_1) (Fig. 3a, b). Total protein was also higher in Gr-1$^+$ cells and bone marrow cells in

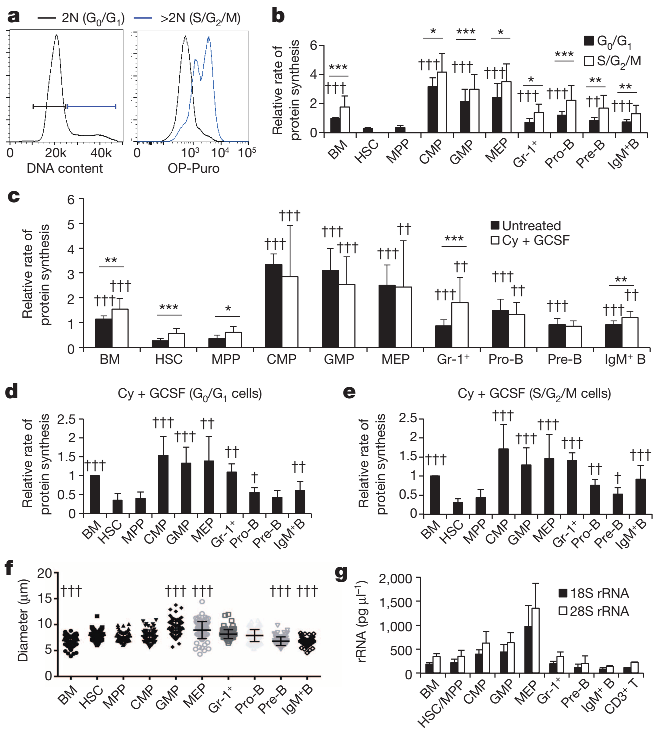


Figure 3 | HSCs synthesize less protein than most haematopoietic progenitors, even when undergoing self-renewing divisions. **a**, OP-Puro incorporation *in vivo* in bone marrow cells in G_0/G_1 versus $S/G_2/M$. 2N, diploid. **b**, Protein synthesis in G_0/G_1 and $S/G_2/M$ cells from haematopoietic-cell populations *in vivo* ($n = 10$ mice from 6 experiments). We were unable to assess OP-Puro incorporation in $S/G_2/M$ HSCs and MPPs in these experiments because these cells are very rare in normal bone marrow. **c**, Protein synthesis in haematopoietic cells after treatment with cyclophosphamide (Cy) and GCSF ($n = 10$ mice from 6 experiments). **d**, **e**, Protein synthesis in G_0/G_1 (**d**) and $S/G_2/M$ (**e**) cells from mice treated with Cy and GCSF ($n = 10$ mice from 6 experiments). Extended Data Fig. 3f, g shows the data from Fig. 3d, e side-by-side with data from untreated controls in Fig. 3b. **f**, Cell diameter ($n > 60$ cells per population from 2 mice). **g**, 18S rRNA and 28S rRNA content in 15,000 cells from each stem- or progenitor-cell population ($n = 3$ mice). All data represent mean ± s.d. To assess the statistical significance of treatment effects within the same cells (**b**, **c**), we performed two-tailed Student's *t*-tests; $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Differences between HSCs and other cells (**b**–**g**) were assessed with a repeated-measures one-way ANOVA followed by Dunnett's test for multiple comparisons; $^\dagger P < 0.05$, $^{\dagger\dagger}P < 0.01$, $^{\dagger\dagger\dagger}P < 0.001$.

$S/G_2/M$ than in G_0/G_1 (Extended Data Fig. 3b). However, HSCs and MPPs exhibited less OP-Puro incorporation than restricted progenitors even when we compared only cells in G_0/G_1 (Fig. 3b). We treated mice with cyclophosphamide followed by two daily injections of granulocyte colony-stimulating factor (GCSF) to induce self-renewing divisions by HSCs (Extended Data Fig. 3d)[25]. Cyclophosphamide and GCSF also increased division by MPPs, Gr-1$^+$ cells and IgM$^+$ B cells (Extended Data Fig. 3a, e). Each of these populations showed increased OP-Puro incorporation after cyclophosphamide and GCSF treatment (Fig. 3c). However, HSCs had significantly less protein synthesis compared to most restricted progenitors, irrespective of whether we compared $S/G_2/M$ cells (Fig. 3e) or G_0/G_1 cells (Fig. 3d and Extended Data Fig. 3f, g) from mice treated with cyclophosphamide and GCSF.

Differences in protein synthesis between HSCs and restricted progenitors were not fully explained by differences in cell diameter (Fig. 3f and Extended Data Fig. 4b), ribosomal RNA (Fig. 3g and Extended Data Fig. 4c, d), or total RNA content (Extended Data Figs 3c and 4e), which were similar among HSCs and lymphoid progenitors despite differences in protein synthesis.

Ribosomal mutant impairs HSC function

Rpl24$^{Bst/+}$ mice have a hypomorphic mutation in the Rpl24 ribosome subunit, reducing protein synthesis in multiple cell types by 30% in culture[3,4,26]. Adult *Rpl24*$^{Bst/+}$ mice show relatively mild phenotypes: they are 20% smaller than wild-type mice and have mild pigmentation and skeletal abnormalities[26]. These *Rpl24*$^{Bst/+}$ mice had normal bone marrow, spleen and thymus cellularity, blood cell counts (Extended Data Fig. 5a, b) and HSC frequency (Fig. 4a). Frequencies of colony-forming progenitors (Extended Data Fig. 6b), restricted progenitors, and annexin V$^+$ HSCs and MPPs (Extended Data Fig. 5c–i) were also largely normal.

OP-Puro incorporation into unfractionated bone marrow cells, HSCs, GMPs and pre-B cells from *Rpl24*$^{Bst/+}$ mice was significantly reduced relative to wild-type cells (Fig. 4d). In the case of HSCs from *Rpl24*$^{Bst/+}$ mice, OP-Puro incorporation was reduced by approximately 30% relative to control HSCs. However, Rpl24 was highly differentially expressed

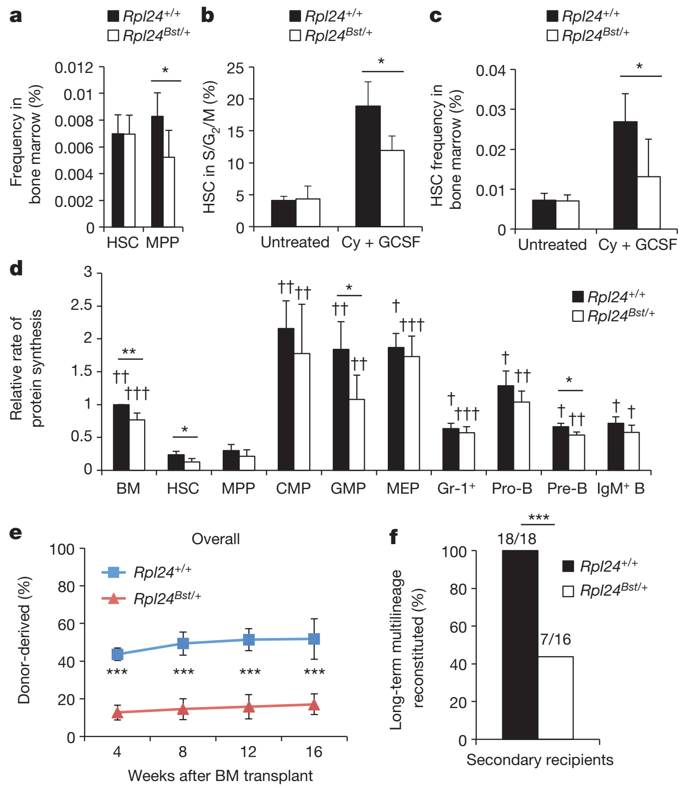


Figure 4 | *Rpl24*$^{Bst/+}$ HSCs synthesize less protein and have less capacity to reconstitute irradiated mice. **a**, Frequencies of HSCs and MPPs in *Rpl24*$^{Bst/+}$ ($n = 7$) versus littermate control ($n = 6$) mice ($n = 5$ experiments). **b**, **c**, Frequency of HSCs in $S/G_2/M$ (**b**; $n = 4$ untreated mice per genotype, $n = 4$ wild-type and 3 *Rpl24*$^{Bst/+}$ mice treated with Cy and GCSF in 3 experiments) and frequency of HSCs in the bone marrow (**c**; $n = 6$ wild-type and 7 *Rpl24*$^{Bst/+}$ untreated mice, $n = 7$ wild-type and 5 *Rpl24*$^{Bst/+}$ Cy-and GCSF-treated mice in 4 experiments) after treatment with Cy and GCSF. **d**, Protein synthesis in haematopoietic cells based on OP-Puro incorporation *in vivo* ($n = 4$ mice per genotype in 4 experiments). **e**, Donor-cell engraftment when 5×10^5 donor bone marrow cells were transplanted along with 5×10^5 recipient bone marrow cells into irradiated recipient mice ($n = 17$ recipient mice of wild-type bone marrow cells and 20 recipient mice of *Rpl24*$^{Bst/+}$ bone marrow cells in 4 experiments; myeloid-, B- and T-cell engraftment are shown in Extended Data Fig. 5k). **f**, The percentage of secondary recipients with long-term multilineage reconstitution (>0.5% donor myeloid and lymphoid cells for at least 16 weeks after transplantation) after secondary transplantation of 3×10^6 bone marrow cells from primary recipients in **e** ($n = 4$ donors per genotype). All data represent mean ± s.d. Statistical significance was assessed with two-tailed Student's *t*-tests (**a**–**e**) and Fisher's exact test (**f**); $^*P < 0.01$...

among haematopoietic cells and some cell populations seemed to depend more than others on Rpl24 for protein synthesis (Extended Data Fig. 5j).

After transplantation into irradiated mice, *Rpl24*$^{Bst/+}$ bone marrow cells gave significantly lower levels of donor cell reconstitution in the myeloid-, B- and T-cell lineages relative to control donor cells (Fig. 4e and Extended Data Fig. 5k). We did not detect impaired homing of *Rpl24*$^{Bst/+}$ LSK cells to the bone marrow (Extended Data Fig. 5l). Significantly impaired reconstitution by *Rpl24*$^{Bst/+}$ cells was also evident in secondary recipient mice (Fig. 4f). After cyclophosphamide and GCSF treatment, significantly fewer *Rpl24*$^{Bst/+}$ HSCs were in S/G_2/M than wild-type HSCs (Fig. 4b). HSC frequency in the bone marrow after cyclophosphamide and GCSF treatment was two- to threefold higher in wild-type than *Rpl24*$^{Bst/+}$ mice (Fig. 4c). Furthermore, colonies formed by individual *Rpl24*$^{Bst/+}$ HSCs in methylcellulose contained significantly fewer cells than wild-type HSC colonies (Extended Data Fig. 6c). *Rpl24*$^{Bst/+}$ HSCs are thus impaired in their proliferative potential.

Some ribosomal defects can induce p53 activation and expression of its target p21^{Cip1} (also known as Cdkn1a)[27]. When we compared *Rpl24*$^{Bst/+}$ with wild-type cells, we did not detect any difference in p53 levels among Lineage^{-} haematopoietic progenitors (Extended Data Fig. 5m) or in p21^{Cip1} levels among LSK cells (Extended Data Fig. 5n). These data are consistent with previous studies[28] indicating that p53 and p21^{Cip1} are not induced in adult *Rpl24*$^{Bst/+}$ haematopoietic cells. Moreover, loss of a single allele of *p53* did not rescue the size of colonies formed by *Rpl24*$^{Bst/+}$ HSCs in culture (Extended Data Fig. 6e), even though p53 heterozygosity largely rescues developmental phenotypes in *Rpl24*$^{Bst/+}$ embryos[28]. HSC defects in *Rpl24*$^{Bst/+}$ mice are therefore not caused by increased p53 function.

Pten deletion increases protein synthesis

HSCs had phosphorylated Akt (pAkt) and pS6 levels that were similar to lymphoid progenitors but lower than myeloid progenitors (Fig. 5a and Extended Data Fig. 4f). We conditionally deleted *Pten* from adult haematopoietic cells in *Mx1-Cre; Pten*$^{fl/fl}$ mice. As shown previously[29–33], *Pten* deletion strongly increased pAKT and pS6 levels in bone marrow cells (Fig. 5a) and in HSCs and MPPs (Fig. 5d). Consistent with this, we observed an approximately 30% increase in protein synthesis in *Pten*-deficient relative to control HSCs ($P < 0.01$; Fig. 5b,c).

HSCs are depleted after *Pten* deletion, even after transplantation into wild-type recipients that never develop leukaemia, by a mechanism that depends on cell-autonomous mTORC1 and mTORC2 activation[29–32]. We examined OP-Puro incorporation into HSCs from *Mx1-Cre; Rictor*$^{fl/fl}$ mice, *Mx1-Cre; Rictor*$^{fl/fl}$; *Pten*$^{fl/fl}$ mice, and *Mx1-Cre; Pten*$^{fl/fl}$ mice. *Rictor* deletion (which inactivates mTORC2) had no effect on the rate of protein synthesis in otherwise wild-type HSCs (Fig. 5c and Extended Data Fig. 6g), consistent with the normal reconstituting capacity of *Rictor*-deficient HSCs[30]. However, *Rictor* deficiency significantly reduced protein synthesis in *Pten*-deficient HSCs to normal levels (Fig. 5c and Extended Data Fig. 6g). The ability of *Rictor* deletion to rescue both protein synthesis and HSC function after *Pten* deletion[30] suggested that *Pten* deletion depletes HSCs partly by cell-autonomously increasing protein synthesis. Rapamycin treatment, which also prevents HSC depletion after *Pten* deletion[31,32], also blocked the increase in protein synthesis in *Pten*-deficient HSCs (data not shown).

HSCs from *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ mice had significantly less protein synthesis than HSCs from *Mx1-Cre; Pten*$^{fl/fl}$ mice (Fig. 5b, c). Although ribosomes can promote mTORC2 signalling[34], *Rpl24*$^{Bst/+}$ did not reduce levels of pAKT, pGSK3β, pS6, or p4EBP1 in HSCs and MPPs (Fig. 5d), suggesting *Rpl24*$^{Bst/+}$ did not reduce mTORC1 or mTORC2 signalling. In fact, *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ HSCs and MPPs had slightly increased levels of pS6 and p4EBP1 relative to *Pten*-deficient HSCs and MPPs, but this would be expected to further increase protein synthesis rather than reducing it. This suggested that *Rpl24*$^{Bst/+}$ acted downstream of the PI3-kinase pathway to block the increase in protein synthesis in HSCs after *Pten* deletion, raising the question of whether *Rpl24*$^{Bst/+}$ could also block leukaemogenesis or HSC depletion.

Rpl24^{Bst} suppresses leukaemogenesis

Conditional deletion of *Pten* in haematopoietic cells leads to a myeloproliferative disorder (MPD) and T-cell acute lymphoblastic leukaemia (T-ALL)[31,33,35]. Increased protein synthesis promotes some haematopoietic malignancies, including T-cell leukaemias[3,4,6,36]. Two weeks after polyinosinic-polycytidylic acid (pIpC) administration, *Mx1-Cre; Pten*$^{fl/fl}$ mice showed significantly increased spleen and thymus cellularity (Fig. 5e, f and Extended Data Fig. 6h, i), consistent with the induction of MPD and T-ALL[31,33,35]. *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ mice exhibited normal spleen and thymus cellularity (Fig. 5e, f and Extended Data Fig. 6h, i), suggesting that reduced ribosome function impaired the development of MPD and T-ALL after *Pten* deletion.

To compare MPD and T-ALL development we transplanted 2×10^6 bone marrow cells from each genetic background into irradiated mice then treated with pIpC 4 weeks later. Recipients of *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ haematopoietic cells lived significantly longer than recipients of *Mx1-Cre; Pten*$^{fl/fl}$ haematopoietic cells (Fig. 5g). We confirmed by PCR that the donor cells had completely excised *Pten* (data not shown). When ill recipients of cells of either genotype were killed, they showed splenomegaly, thymomegaly and histological signs of MPD and T-ALL (Extended Data Fig. 6k). *Rpl24*$^{Bst/+}$ therefore significantly delayed ($P < 0.001$; Fig. 5g), but did not entirely prevent, MPD and T-ALL after *Pten* deletion. It is unclear whether *Rpl24*$^{Bst/+}$ impairs leukaemogenesis by acting within HSCs or within other haematopoietic cells.

Pten deletion mobilizes HSCs to the spleen[29,30]. *Rpl24*$^{Bst/+}$ blocked this effect, restoring normal HSC numbers in spleens of *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ mice (Extended Data Fig. 6j).

Rpl24^{Bst} rescues *Pten*-deficient HSCs

We performed long-term reconstitution assays in which we transplanted 10 CD150^{+}CD48^{-}LSK HSCs from wild-type, *Mx1-Cre; Pten*$^{fl/fl}$, *Rpl24*$^{Bst/+}$, or *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ mice along with 300,000 wild-type bone marrow cells into irradiated wild-type mice. As expected, most (15 of 18) recipients of wild-type HSCs but no (0 of 12) recipients of *Pten*-deficient HSCs showed long-term multilineage reconstitution by donor cells (Fig. 5h, i). Relative to control HSCs, *Rpl24*$^{Bst/+}$ HSCs gave lower levels of donor-cell reconstitution in all lineages (Fig. 5h) and a somewhat lower fraction of recipients showed long-term multilineage reconstitution (10 of 19; Fig. 5i), although these differences were not statistically significant. Relative to *Pten*-deficient HSCs, *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ compound mutant HSCs gave significantly higher levels of reconstitution and a significantly higher percentage of recipients showed long-term multilineage reconstitution by donor cells (Fig. 5h, i). Reconstitution by *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ HSCs was statistically indistinguishable from wild-type HSCs in all lineages except the B lineage (*Pten* deficiency impairs B-lineage progenitors independent of its effects on HSCs[33]; Fig. 5h). We confirmed by PCR that the donor cells had completely excised *Pten* (data not shown). This demonstrates that reducing protein synthesis restores the ability of *Pten*-deficient HSCs to give long-term multilineage reconstitution.

Secondary transplantation of 3×10^6 bone marrow cells from primary recipients with long-term multilineage reconstitution in Fig. 5h produced significantly less donor cell reconstitution and long-term multilineage reconstitution among recipients of *Rpl24*$^{Bst/+}$ compared to control HSCs (Fig. 5j, k). Secondary recipients of *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ compound mutant HSCs also showed significantly less donor cell reconstitution and long-term multilineage reconstitution compared to control HSCs. *Rpl24*$^{Bst/+}$ thus restored long-term multilineage reconstituting potential to *Pten*-deficient HSCs but did not fully restore wild-type function. *Pten* deficiency thus impairs HSC function, in part by increasing protein synthesis.

To test MPP function we competitively transplanted 100 donor CD150^{-}CD48^{-}LSK MPPs[19] from *Mx-1-Cre; Pten*$^{fl/fl}$, *Rpl24*$^{Bst/+}$, or control mice into irradiated recipients. There was a trend towards lower reconstitution by *Rpl24*$^{Bst/+}$ MPPs and *Pten*-deficient MPPs relative to control MPPs but the differences were not statistically significant

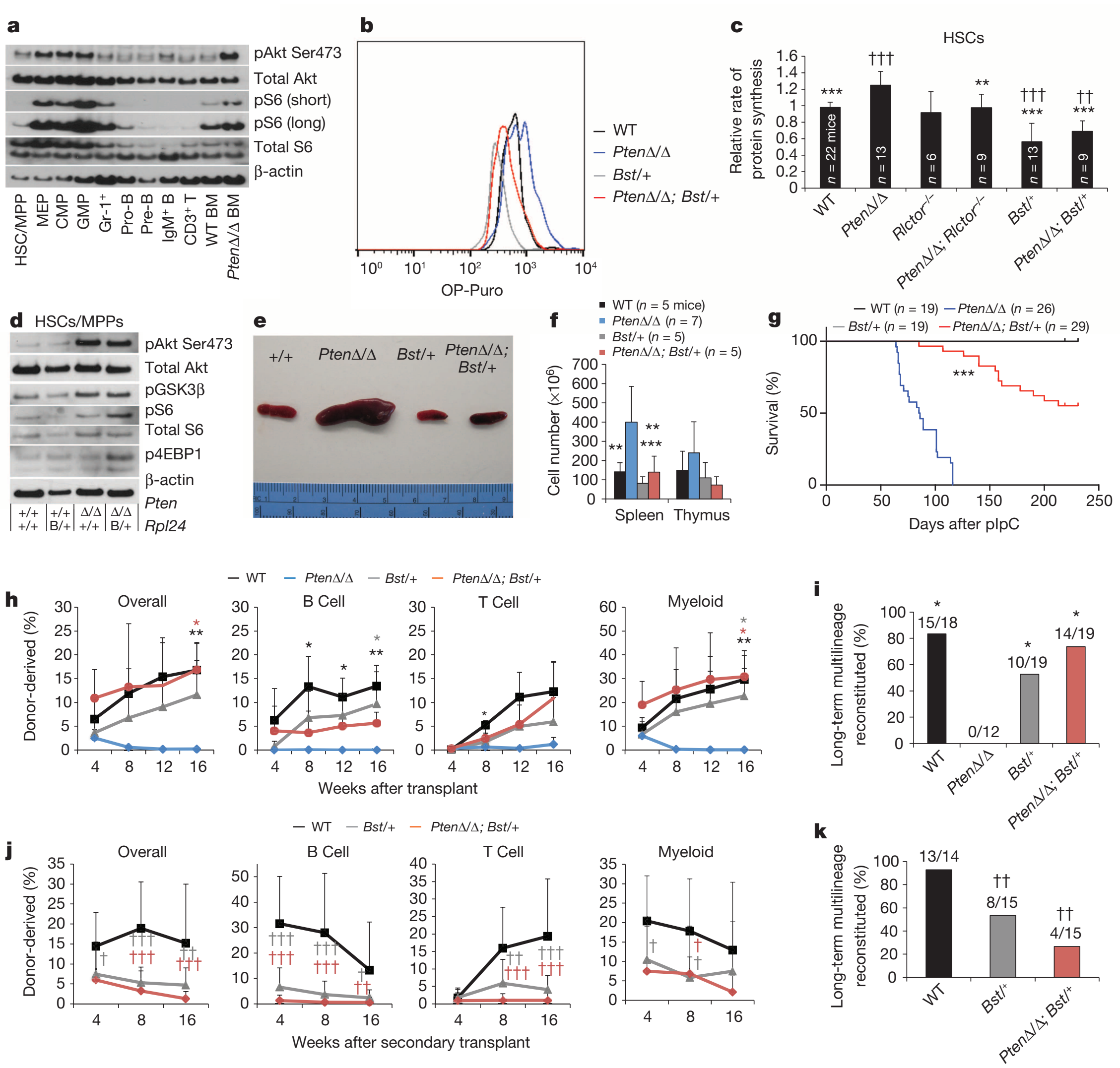


Figure 5 | *Rpl24*$^{Bst/+}$ blocks the increase in protein synthesis and restores HSC function after *Pten* deletion. **a**, Western blot analyses of 30,000 cells from each population. Long and short exposures for pS6 are shown. For total S6, a non-specific band is present below the specific band. Differences in β-actin represent differences in β-actin content per cell (one representative blot from two experiments). WT, wild type. **b**, Representative histograms of OP-Puro fluorescence in HSCs of each genotype. **c**, OP-Puro incorporation into HSCs of each genotype ($n = 15$ experiments). **d**, Western blot analyses of 30,000 CD48^{-}LSK cells (HSCs and MPPs) of the indicated genotypes (one representative blot from two experiments). **e**, Representative spleens 2 weeks after pIpC administration to wild-type, *Mx-1-Cre; Pten*$^{fl/fl}$ (*Pten*Δ/Δ), *Rpl24*$^{Bst/+}$ (*Bst*/+), and *Mx-1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ (*Pten*Δ/Δ; *Bst*/+) mice. **f**, Spleen and thymus cellularity ($n = 7$ experiments). **g**, Time until mice had to be euthanized owing to illness after transplantation of 2×10^6 bone marrow cells of the indicated genotypes into irradiated recipient mice. **h**, **i**, Ten donor HSCs were transplanted along with 3×10^5 recipient bone marrow cells into irradiated recipients. Donor-cell engraftment (**h**) and fraction of recipients that showed long-term multilineage reconstitution (**i**; 3 experiments). **j**, **k**, Donor-cell engraftment (**j**) and the fraction of secondary recipients that showed long-term multilineage reconstitution (**k**) after transplantation of 3×10^6 bone marrow cells from primary recipients in **h** ($n = 4$ donors per genotype). All data represent mean ± s.d. Differences among genotypes (**c**, **f**, **h**) were assessed with a one-way ANOVA followed by Dunnett's test for multiple comparisons. Statistical significance was assessed by log-rank test (**g**), chi-squared tests followed by Tukey's *t*-tests for pairwise comparisons (**i**, **k**), or a one-way ANOVA followed by Tukey's *t*-tests for multiple comparisons (**j**). Significance was expressed relative to wild-type (†$P < 0.05$, ††$P < 0.01$, †††$P < 0.001$) or *Pten*-deficient cells (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

(Extended Data Fig. 7a, b). Neither *Rpl24*$^{Bst/+}$ nor *Pten* deletion significantly affected the percentage of bone marrow cells that formed colonies or the cellularity of those colonies (Extended Data Fig. 6b, d). Thus, we have not detected clear effects of *Rpl24*$^{Bst/+}$ or *Pten* deletion on the proliferative potential of MPPs or restricted progenitors, suggesting that they are not as sensitive as HSCs to changes in protein synthesis, although it remains possible that they are also impaired by changes in protein synthesis.

DISCUSSION

When we added HSCs to culture, OP-Puro incorporation increased dramatically (data not shown). This raises the possibility that the failure to maintain HSCs sustainably under any known culture conditions[37] may reflect a limitation imposed by increased protein synthesis. Consistent with this, rapamycin promotes HSC maintenance in culture[38], although it remains to be determined whether rapamycin attenuates the increase in protein synthesis in cultured HSCs.

Low rates of protein synthesis may be essential for maintaining metabolic homeostasis in HSCs and potentially in other kinds of somatic stem cells. Changes in protein synthesis may cause undesirable changes in the quality and/or content of the proteome, such as due to misfolding. There may also be changes in the translation of certain subsets of transcripts (potentially including key HSC regulators) when protein synthesis increases, similar to what occurs in cancer cells[5].

Embryonic stem cells have high proteasome activity and proteasome activity increases both embryonic-stem-cell maintenance and *Caenorhabditis elegans* lifespan[39,40]. Rapamycin treatment, and mutations that reduce mTOR signalling, also increase lifespan (see citations in ref. 41) and would be predicted to reduce protein synthesis. In the context of these results, our observations raise the possibility that reduced protein synthesis and/or increased proteasome activity is required to maintain certain long-lived somatic cells in addition to increasing organismal lifespan.

METHODS SUMMARY

For *in vivo* quantification of protein synthesis, OP-Puro (50 mg kg^{-1}; pH 6.4–6.6 in PBS; Medchem Source) was injected intraperitoneally. Unless otherwise indicated, bone marrow was collected 1 h later and 3×10^6 cells were stained with antibodies against cell surface markers, fixed in 1% paraformaldehyde, and permeabilized in PBS with 3% fetal bovine serum and 0.1% saponin, as described in the Methods. OP-Puro was detected by performing an azide-alkyne cycloaddition with the Click-iT Cell Reaction Buffer Kit (Life Technologies), and analysed by flow cytometry as described in the Methods. 'Relative rates of protein synthesis' were calculated by normalizing OP-Puro signals to bone marrow after subtracting background fluorescence. 'Mean OP-Puro fluorescence' reflected absolute fluorescence values for each cell population from multiple experiments.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 May 2013; accepted 15 January 2014.
Published online 9 March 2014.

1. Narla, A. & Ebert, B. L. Ribosomopathies: human disorders of ribosome dysfunction. *Blood* **115,** 3196–3205 (2010).
2. Sakamoto, K. M., Shimamura, A. & Davies, S. M. Congenital disorders of ribosome biogenesis and bone marrow failure. *Biol. Blood Marrow Transplant.* **16,** S12–S17 (2010).
3. Hsieh, A. C. *et al.* Genetic dissection of the oncogenic mTOR pathway reveals druggable addiction to translational control via 4EBP-eIF4E. *Cancer Cell* **17,** 249–261 (2010).
4. Barna, M. *et al.* Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature* **456,** 971–975 (2008).
5. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485,** 55–61 (2012).
6. Ruggero, D. & Pandolfi, P. P. Does the ribosome translate cancer? *Nature Rev. Cancer* **3,** 179–192 (2003).
7. Jaako, P. *et al.* Mice with ribosomal protein S19 deficiency develop bone marrow failure and symptoms like patients with Diamond–Blackfan anemia. *Blood* **118,** 6087–6096 (2011).
8. Wong, C. C., Traynor, D., Basse, N., Kay, R. R. & Warren, A. J. Defective ribosome assembly in Shwachman–Diamond syndrome. *Blood* **118,** 4305–4312 (2011).
9. Danilova, N., Sakamoto, K. M. & Lin, S. Ribosomal protein S19 deficiency in zebrafish leads to developmental abnormalities and defective erythropoiesis through activation of p53 protein family. *Blood* **112,** 5228–5237 (2008).
10. Sen, S. *et al.* The ribosome-related protein, SBDS, is critical for normal erythropoiesis. *Blood* **118,** 6407–6417 (2011).
11. Payne, E. M. *et al.* L–Leucine improves the anemia and developmental defects associated with Diamond-Blackfan anemia and del(5q) MDS by activating the mTOR pathway. *Blood* **120,** 2214–2224 (2012).
12. Beatty, K. E. *et al.* Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem. Int. Edn Engl.* **45,** 7364–7367 (2006).
13. Nathans, D. Puromycin inhibition of protein synthesis: incorporation of puromycin into peptide chains. *Proc. Natl Acad. Sci. USA* **51,** 585–592 (1964).
14. Schmidt, E. K., Clavarino, G., Ceppi, M. & Pierre, P. SUnSET, a nonradioactive method to monitor protein synthesis. *Nature Methods* **6,** 275–277 (2009).
15. Starck, S. R., Green, H. M., Alberola-Ila, J. & Roberts, R. W. A general approach to detect protein expression *in vivo* using fluorescent puromycin conjugates. *Chem. Biol.* **11,** 999–1008 (2004).
16. Joseph, N. M. & Morrison, S. J. Toward and understanding of the physiological function of mammalian stem cells. *Dev. Cell* 173–183 (2005).
17. Liu, J., Xu, Y., Stoleru, D. & Salic, A. Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. *Proc. Natl Acad. Sci. USA* **109,** 413–418 (2012).
18. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121,** 1109–1121 (2005).
19. Oguro, H., Ding, L. & Morrison, S. J. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell* **13,** 102–116 (2013).
20. Zhou, S. *et al. Bcrp1* gene expression is required for normal numbers of side population stem cells in mice, and confers relative protection to mitoxantrone in hematopoietic cells *in vivo. Proc. Natl Acad. Sci. USA* **99,** 12339–12344 (2002).
21. Ciechanover, A., Finley, D. & Varshavsky, A. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. *Cell* **37,** 57–66 (1984).
22. Luker, G. D., Pica, C. M., Song, J., Luker, K. E. & Piwnica-Worms, D. Imaging 26S proteasome activity and inhibition in living mice. *Nature Med.* **9,** 969–973 (2003).
23. Gingras, A. C., Raught, B. & Sonenberg, N. Regulation of translation initiation by FRAP/mTOR. *Genes Dev.* **15,** 807–826 (2001).
24. Wek, R. C., Jiang, H. Y. & Anthony, T. G. Coping with stress: eIF2 kinases and translational control. *Biochem. Soc. Trans.* **34,** 7–11 (2006).
25. Morrison, S. J., Wright, D. & Weissman, I. L. Cyclophosphamide/granulocyte colony-stimulating factor induces hematopoietic stem cells to proliferate prior to mobilization. *Proc. Natl Acad. Sci. USA* **94,** 1908–1913 (1997).
26. Oliver, E. R., Saunders, T. L., Tarle, S. A. & Glaser, T. Ribosomal protein L24 defect in belly spot and tail (*Bst*), a mouse *Minute. Development* **131,** 3907–3920 (2004).
27. Fumagalli, S. & Thomas, G. The role of p53 in ribosomopathies. *Semin. Hematol.* **48,** 97–105 (2011).
28. Barkić, M. *et al.* The p53 tumor suppressor causes congenital malformations in *Rpl24*-deficient mice and promotes their survival. *Mol. Cell. Biol.* **29,** 2489–2504 (2009).
29. Lee, J. Y. *et al.* mTOR activation induces tumor suppressors that inhibit leukemogenesis and deplete hematopoietic stem cells after *Pten* deletion. *Cell Stem Cell* **7,** 593–605 (2010).
30. Magee, J. A. *et al.* Temporal changes in PTEN and mTORC2 regulation of hematopoietic stem cell self-renewal and leukemia suppression. *Cell Stem Cell* **11,** 415–428 (2012).
31. Yilmaz, O. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441,** 475–482 (2006).
32. Kalaitzidis, D. *et al.* mTOR Complex 1 plays critical roles in hematopoiesis and pten-loss-evoked leukemogenesis. *Cell Stem Cell* **11,** 429–439 (2012).
33. Zhang, J. *et al.* PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* **441,** 518–522 (2006).
34. Zinzalla, V., Stracka, D., Oppliger, W. & Hall, M. N. Activation of mTORC2 by association with the ribosome. *Cell* **144,** 757–768 (2011).
35. Guo, W. *et al.* Multi-genetic events collaboratively contribute to *Pten*-null leukaemia stem-cell formation. *Nature* **453,** 529–533 (2008).
36. Ruggero, D. Translational control in cancer etiology. *Cold Spring Harb. Perspect. Biol.* **5,** a012336 (2013).
37. Dahlberg, A., Delaney, C. & Bernstein, I. D. *Ex vivo* expansion of human hematopoietic stem and progenitor cells. *Blood* **117,** 6083–6090 (2011).
38. Huang, J., Nguyen-McCarty, M., Hexner, E. O., Danet-Desnoyers, G. & Klein, P. S. Maintenance of hematopoietic stem cells through regulation of Wnt and mTOR pathways. *Nature Med.* **18,** 1778–1785 (2012).
39. Vilchez, D. *et al.* Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. *Nature* **489,** 304–308 (2012).
40. Vilchez, D. *et al.* RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. *Nature* **489,** 263–268 (2012).
41. Signer, R. A. & Morrison, S. J. Mechanisms that regulate stem cell aging and life span. *Cell Stem Cell* **12,** 152–165 (2013).

Acknowledgements S.J.M. is a Howard Hughes Medical Institute Investigator, the Mary McDermott Cook Chair in Pediatric Genetics, and the director of the Hamon Laboratory for Stem Cells and Cancer. This work was supported by the Cancer Prevention and Research Institute of Texas and the National Institute on Aging (R37 AG024945). R.A.J.S. was supported by fellowships from the Leukemia & Lymphoma Society (5541-11) and the Canadian Institutes of Health Research (MFE-106993). J.A.M. was supported by the UT Southwestern K12 Pediatrics Training Grant (K12-HD068369). We thank A. Pineda, K. Cowan, E. Daniel, M. Acar, H. Oguro, J. Peyer, K. Rajagopalan, M. Agathocleous and E. Piskounova for technical support and advice; N. Loof and the Moody Foundation Flow Cytometry Facility, L. Hynan and J. Reisch for advice regarding statistics; J. Shelton for histology; and R. Coolon, S. Manning, M. Gross and K. Correll for mouse colony management.

Author Contributions R.A.J.S. conceived the project. A.S. developed the OP-Puro reagent and provided advice in the early stages of the project. R.A.J.S. performed all of the experiments with the exception of the western blot analyses in Figs 2e and 5a, d, and Extended Data Fig. 5j, which were performed by J.A.M. R.A.J.S., J.A.M. and S.J.M. designed the experiments and interpreted results. R.A.J.S. and S.J.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.M. (sean.morrison@utsouthwestern.edu).

METHODS

Mice. *Rpl24*$^{Bst/+}$ (ref. 26), *Pten*$^{fl/fl}$ (ref. 42), *Mx1-Cre* (ref. 43), *Rictor*$^{fl/fl}$ (ref. 30), *Abcg2*$^{-/-}$ (ref. 44) and *p53*$^{+/-}$ (ref. 45) mutant mice have been described previously. These mice were all backcrossed for at least eight generations onto a C57BL background, with the exception of *Abcg2*$^{-/-}$ mice, which were on an FVB.129 N7 background (Taconic). C57BL/Ka-Thy-1.1 (CD45.2) and C57BL/Ka-Thy-1.2 (CD45.1) mice were used in transplantation experiments. Both male and female mice were used in all studies. Expression of *Mx1-Cre* was induced by three or four intraperitoneal injections of 10 μg pIpC (GE Healthcare) administered every other day, beginning at approximately 6 weeks of age. For cyclophosphamide and GCSF experiments, 4 mg of cyclophosphamide (Baxter) was administered by intraperitoneal injection on day 0, and 5 μg of GCSF (Neupogen; Amgen) was administered by subcutaneous injection on days 1 and 2. Mice were analysed on day 3. All mice were housed in the Unit for Laboratory Animal Medicine at the University of Michigan, where breeding for these studies was initiated, or in the Animal Resource Center at the University of Texas Southwestern Medical Center, where these studies were performed. All protocols were approved by the University of Michigan Committee on the Use and Care of Animals and by the University of Texas Southwestern Institutional Animal Care and Use Committee.

Measurement of protein synthesis. For *in vitro* analysis 10^3–10^4 bone marrow or sorted cells were plated in 100 μl of methionine-free Dulbecco's Modified Eagle's Medium (Sigma) supplemented with 200 μM L-cysteine (Sigma), 50 μM 2-mercaptoethanol (Sigma), 1 mM L-glutamine (Gibco) and 0.1% bovine serum albumin (BSA; Sigma). For analysis of HPG and AHA incorporation, cells were pre-cultured for 45 min to deplete endogenous methionine. For OP-Puro, the medium was supplemented with 1 mM L-methionine (Sigma). HPG (Life Technologies; 1 mM final concentration), AHA (Life Technologies; 1 mM final concentration) or OP-Puro (Medchem Source; 50 μM final concentration) were added to the culture medium for 1 h (HPG and OP-Puro) or 2.5 h (AHA), then cells were removed from wells and washed twice in Ca^{2+}- and Mg^{2+}-free phosphate buffered saline (PBS). Cells were fixed in 0.5 ml of 1% paraformaldehyde (Affymetrix) in PBS for 15 min on ice. Cells were washed in PBS, then permeabilized in 200 μl PBS supplemented with 3% fetal bovine serum (Sigma) and 0.1% saponin (Sigma) for 5 min at room temperature (20–25 °C). The azide-alkyne cycloaddition was performed using the Click-iT Cell Reaction Buffer Kit (Life Technologies) and azide conjugated to Alexa Fluor 488 or Alexa Fluor 555 (Life Technologies) at 5 μM final concentration. After the 30-min reaction, the cells were washed twice in PBS supplemented with 3% fetal bovine serum and 0.1% saponin, then resuspended in PBS supplemented with 4′,6-diamidino-2-phenylindole (DAPI; 4 μg ml^{-1} final concentration) and analysed by flow cytometry. To inhibit OP-Puro, HPG or AHA incorporation, cycloheximide (Sigma) was added 30 min before OP-Puro, HPG or AHA at a final concentration of 100 μg ml^{-1}. All cultures were incubated at 37 °C in 6.5% CO_2 and constant humidity.

For *in vivo* analysis, OP-Puro (50 mg kg^{-1} body mass; pH 6.4–6.6 in PBS) was injected intraperitoneally. One hour later mice were euthanized, unless indicated otherwise. Bone marrow was collected, and 3×10^6 cells were stained with combinations of antibodies against cell-surface markers as described below. After washing, the cells were fixed, permeabilized, and the azide-alkyne cycloaddition was performed as described above. 'Relative rates of protein synthesis' were calculated by normalizing OP-Puro signals to whole bone marrow after subtracting autofluorescence background. 'Mean OP-Puro fluorescence' reflected absolute fluorescence values for each cell population from multiple independent experiments.

To assess the effect of proteasome activity on OP-Puro incorporation mice were administered an intravenous injection of bortezomib (Cell Signaling; 1 mg kg^{-1} body mass) 1 h before OP-Puro administration. OP-Puro incorporation was assessed as described above 1 h later unless indicated otherwise.

Flow cytometry and cell isolation. Bone marrow cells were isolated by flushing the long bones (femurs and tibias) or by crushing the long bones, vertebrae and pelvic bones with a mortar and pestle in Ca^{2+}- and Mg^{2+}-free Hank's buffered salt solution (HBSS; Gibco) supplemented with 2% heat-inactivated bovine serum (Gibco). Spleens and thymuses were prepared by crushing tissues between frosted slides. All cells were filtered through a 40-μm cell strainer to obtain single cell suspensions. Cell number and viability were assessed by a Vi-CELL cell viability analyser (Beckman Coulter) or by counting with a hemocytometer.

For flow cytometric analysis and isolation of specific haematopoietic progenitors, cells were incubated with combinations of antibodies to the following cell-surface markers, conjugated to FITC, PE, PerCP-Cy5.5, Cy5, APC, PE-Cy7, eFluor 660, Alexa Fluor 700, APC-eFluor 780 or biotin (antibodies are given in brackets in the following list): CD2 (RM2-5), CD3ε (17A2), CD4 (GK1.5), CD5 (53-7.3), CD8α (53-6.7), CD11b (M1/70), CD16/32 (FcYRII/III; 93), CD24 (M1/69), CD25 (PC61.5), CD34 (RAM34), CD43 (1B11), CD44 (IM7), CD45.1 (A20), CD45.2 (104), CD45R (B220; RA3-6B2), CD48 (HM48-1), CD71 (C2), CD117 (cKit; 2B8), CD127 (IL7Rα; A7R34), CD138 (281-2), CD150 (TC15-12F12.2), Ter119 (TER-119), Sca-1 (D7, E13-161.7), Gr-1 (RB6-8C5) and IgM (II/41). For isolation of HSCs and MPPs, Lineage markers included CD3, CD5, CD8, B220, Gr-1 and Ter119. For isolation of CMPs, GMPs and MEPs, these Lineage markers were supplemented with additional antibodies against CD4 and CD11b.

Biotinylated antibodies were visualized by incubation with PE-Cy7 or APC-Alexa Fluor 750 conjugated streptavidin. All reagents were acquired from BD Biosciences, eBiosciences, or BioLegend. All incubations were for approximately 30 min on ice. HSCs, MPPs, CD34$^+$CD16/32^{low}CD127$^-$Lineage$^-$Sca-1$^-$c-kit$^+$ CMPs (ref. 46), CD34$^+$CD16/32^{high}CD127$^-$Lineage$^-$Sca-1$^-$c-kit$^+$ GMPs (ref. 46), and CD34$^-$CD16/32$^{-/low}$CD127$^-$Lineage$^-$Sca-1$^-$c-kit$^+$ MEPs (ref. 46) were sometimes pre-enriched by selecting c-kit$^+$ cells using paramagnetic microbeads and an autoMACS magnetic separator (Miltenyi Biotec).

Non-viable cells were excluded from sorts and analyses using DAPI (4′,6-diamidino-2-phenylindole). Apoptotic cells were identified using APC annexin V (BD Biosciences). Data acquisition and cell sorting were performed on a FACSAria flow cytometer (BD Biosciences). All sorted fractions were double sorted to ensure high purity. Data were analysed by FACSDiva (BD Biosciences) or FlowJo (Tree Star) software.

Long-term competitive repopulation assay. Adult recipient mice (CD45.1) were administered a minimum lethal dose of radiation using an XRAD 320 X-ray irradiator (Precision X-Ray) to deliver two doses of 540 rad (1,080 rad in total) at least 3 h apart. Cells were injected into the retro-orbital venous sinus of anaesthetized recipients. For competitive bone marrow transplants 5×10^5 donor and 5×10^5 recipient cells were transplanted. For HSC transplants 10 donor CD150$^+$CD48$^-$Lineage$^-$Sca-1$^+$c-kit$^+$ HSCs and 3×10^5 recipient bone marrow cells were transplanted. Blood was obtained from the tail veins of recipient mice every 4 weeks for at least 16 weeks after transplantation. For MPP transplants 100 donor CD150$^-$CD48$^-$Lineage$^-$Sca-1$^+$c-kit$^+$ MPPs and 3×10^5 recipient bone marrow cells were transplanted. Blood was obtained from the tail veins of recipient mice 3, 5 and 7 weeks after transplantation. Red blood cells were lysed with ammonium chloride potassium buffer. The remaining cells were stained with antibodies against CD45.2, CD45.1, CD45R (B220), CD11b, CD3 and Gr-1 to assess donor-cell engraftment. Mice that died or developed obvious haematopoietic neoplasms were omitted from the analyses. For secondary transplants, 3×10^6 bone marrow cells collected from primary recipients were transplanted non-competitively into irradiated recipient mice. Primary recipients used for secondary transplantation had long-term multilineage reconstitution by donor cells and levels of donor-cell reconstitution that were typical (closest to mean values) for the treatments from which they originated. For survival studies in Fig. 5, irradiated mice were transplanted with 2×10^6 bone marrow cells of each genotype and pIpC was administered 4 weeks later to delete *Pten*. Three to four donors per genotype were used with 5 to 10 recipients per donor. Mice were euthanized when they appeared moribund.

Western blot analysis. Equal numbers of cells from each stem or progenitor population (unless indicated otherwise) were sorted into, or resuspended in, trichloracetic acid (TCA, Sigma). The final concentration was adjusted to 10% TCA. Extracts were incubated on ice for at least 15 min and centrifuged at 16,100*g* at 4 °C for 15 min. Precipitates were washed in acetone twice and dried. The pellets were solubilized in 9 M urea, 2% Triton X-100, and 1% DTT. LDS loading buffer (Life Technologies) was added and the pellet was heated at 70 °C for 10 min. Samples were separated on Bis-Trispolyacrylamide gels (Life Technologies) and transferred to PVDF membranes (Millipore). Western blot analyses were performed according to the protocol from Cell Signaling Technologies and blots were developed with the SuperSignal West Femto chemiluminescence kit (Thermo Scientific). Blots were stripped with 1% SDS, 25 mM glycine (pH 2) before re-probing. The following primary antibodies were used for western blot analyses (obtained from Cell Signaling Technologies unless indicated otherwise): phos-Akt (Ser473; D9E), AKT (C67E7), phos-S6 (Ser240/244; polyclonal), S6 (5G10), phos-4E-BP1 (Thr37/46; 236B4), 4E-BP1 (polyclonal), phos-eIF2α (Ser51; polyclonal), eIF2α (D7D3), phos-GSK3β (Ser9; D85E12), β-actin (AC-74; Sigma), p53 (cm5p; Leica), p21^{Cip1} (F-5; Santa Cruz) and Rpl24 (polyclonal; Abcam). For p21^{Cip1}, the membranes were treated with the SuperSignal Western Blot Enhancer (Thermo Scientific). Band intensity was quantified with ImageJ software.

Methylcellulose cultures. Two-hundred live bone marrow cells or single CD150$^+$CD48$^-$LSK cells were sorted per well of a 96-well plate containing methylcellulose culture medium (M3434, StemCell Technologies) and incubated at 37 °C in 6.5% CO_2 and constant humidity. Colony formation was assessed 14 days after plating, except for data shown in Extended Data Fig. 6e, in which colonies were assessed 15 days after plating. Colony size was assessed by picking individual colonies with a pipette under the microscope, washing the cells in PBS, and counting live cells by Trypan blue exclusion on a hemocytometer.

Cell size. The average diameter of sorted cells was measured by sorting cells into flat-bottom 96-well plates and analysing micrographs with ImageJ software.

RNA content. From each haematopoietic-cell population, 15,000 cells were sorted into RLT Plus buffer (Qiagen) supplemented with 2-mercaptoethanol. RNA from each sample was extracted with the RNeasy micro plus kit (Qiagen) into 14 μl of water. Two microlitres of each sample were analysed in duplicate for 18S rRNA, 28S rRNA and total RNA concentration using an Agilent RNA 6000 Pico kit and a Bioanalyzer (Agilent) at the University of Texas Southwestern Genomics and Microarray Core.

Protein content. From each population, 50,000 cells were sorted into PBS. G_0/G_1 cells were distinguished from $S/G_2/M$ cells based on DNA content using Hoechst 33342 (Sigma) staining of live cells. Cells were spun down and resuspended in 20 μl of RIPA buffer (Pierce) supplemented with complete protease inhibitor (Roche) and rotated at 4 °C for 30 min. Lysates were spun down at 16,100*g* at 4 °C for 15 min. The protein content in the supernatant was assessed with the BCA assay (Pierce).

Statistical methods. In all cases, multiple independent experiments were performed on different days to verify the reproducibility of experimental findings. Group data are always represented by mean ± standard deviation. The numbers of experiments noted in figure legends reflect independent experiments performed on different days.

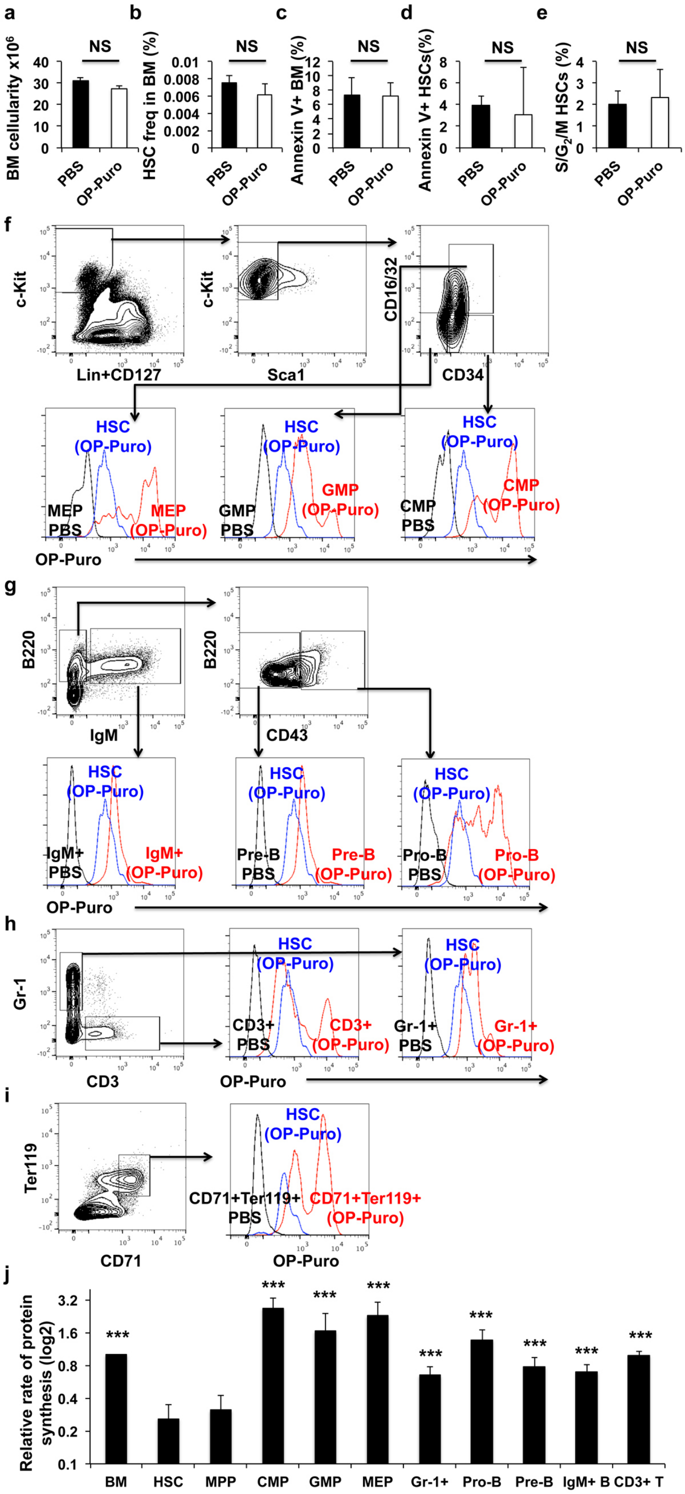
To test statistical significance between two samples, two-tailed Student's *t*-tests were used. When multiple samples were compared, statistical significance was assessed using a one-way ANOVA or a repeated-measures one-way ANOVA (when comparing multiple time points or populations from the same mouse) followed by Dunnett's test for multiple comparisons. When multiple samples were each compared to one another, statistical significance was assessed using a one-way ANOVA followed by Tukey's *t*-tests for multiple comparisons. Statistical significance comparing overall numbers of mice with long-term multilineage reconsitution was assessed either by a Fisher's exact test (Fig. 4f) or by chi-squared tests followed by Tukey's *t*-tests for pairwise comparisons (Fig. 5i, k). Statistical significance with respect to differences in survival (Fig. 5g) was calculated using a log-rank test. The specific type of test used for each figure panel is described in the figure legends.

For normalized protein synthesis rates and normalized messenger RNA expression, means were calculated and statistical tests were performed using $\log_{10}$-transformed data and then means were back-transformed to prevent data skewing.

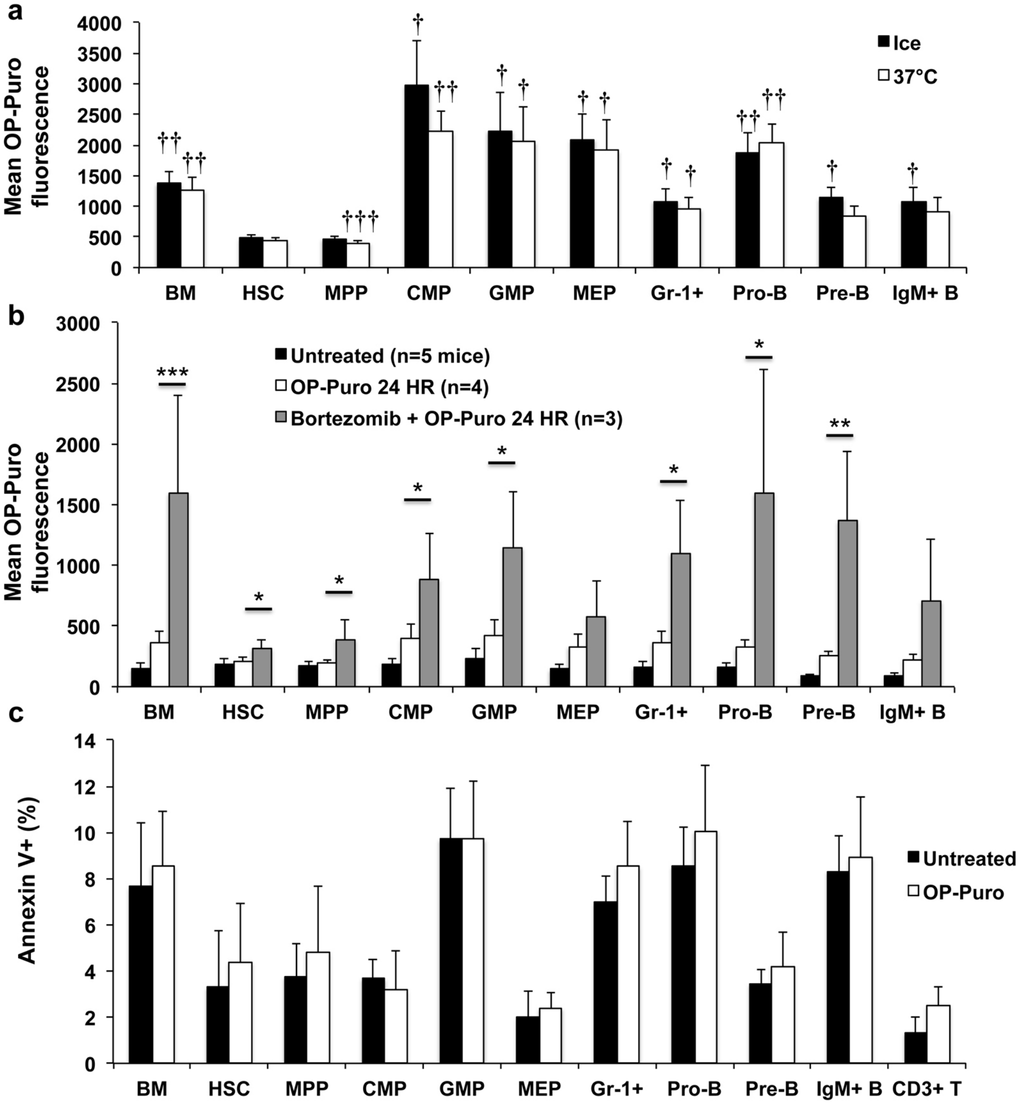
No randomization or blinding was used in any experiments. The only mice excluded from any experiment were those that developed leukaemia after transplantation in Fig. 5h, i. As the purpose of this experiment was to compare the reconstituting capacity of normal HSCs, the presence of leukaemia in a minority of recipient mice was a confounding factor that had the potential to inappropriately skew the results; therefore, the experiment was initiated with the intention of excluding data from any mouse that died during the experiment. We excluded 0 to 4 mice per treatment.

In the case of measurements in which variation among experiments tends to be low (for example, HSC frequency) we generally examined 3 to 6 mice. In the case of measurements in which variation among experiments tends to be higher (for example, reconstitution assays) we examined larger numbers of mice ($>$10). In the case of assays to assess protein synthesis, there were no historical data on which to base sample sizes, and we therefore performed multiple independent experiments with multiple biological replicates to ensure the reproducibility of our findings.

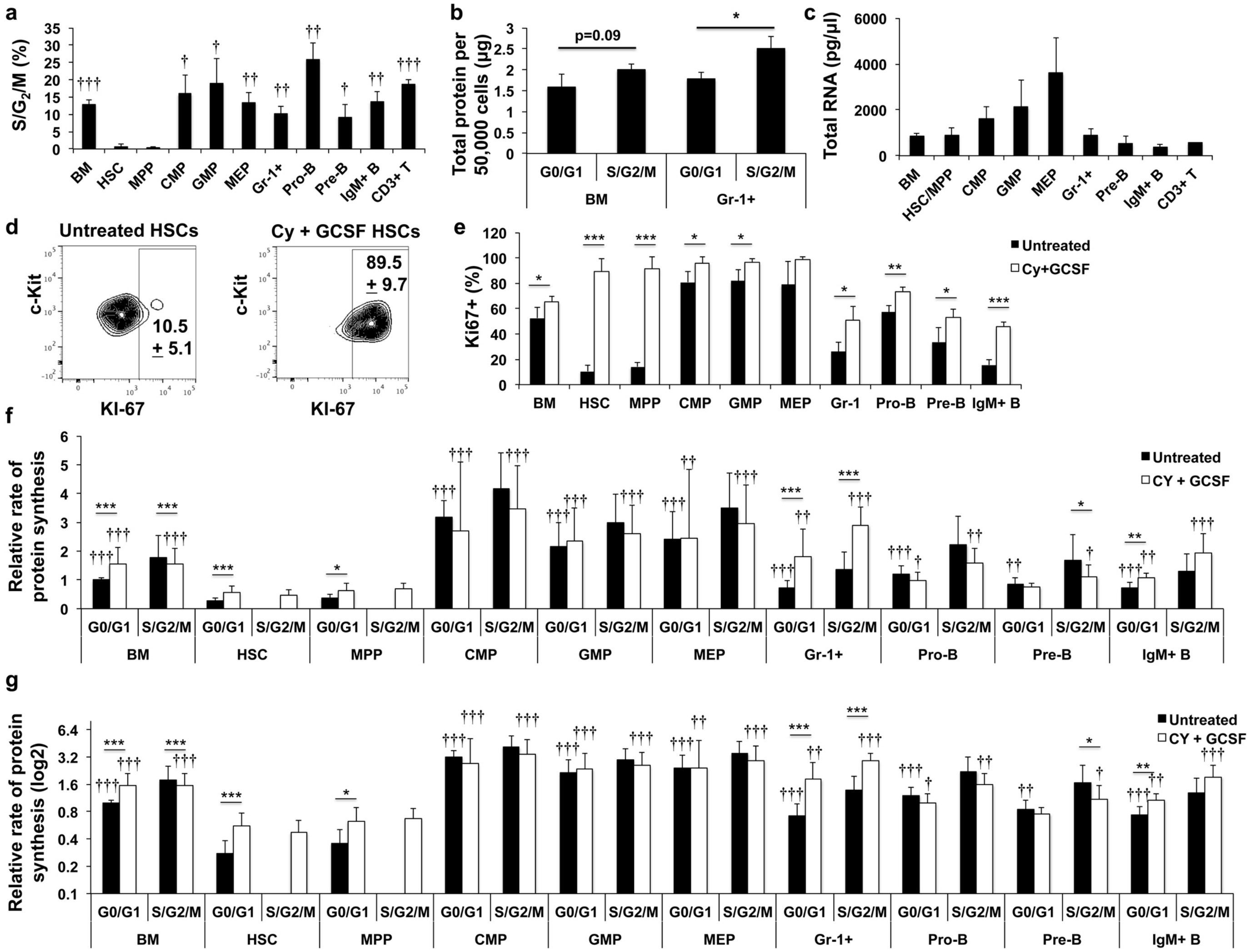
42. Groszer, M. *et al.* PTEN negatively regulates neural stem cell self-renewal by modulating G0-G1 cell cycle entry. *Proc. Natl Acad. Sci. USA* **103,** 111–116 (2006).
43. Kühn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269,** 1427–1429 (1995).
44. Jonker, J. W. *et al.* The breast cancer resistance protein protects against a major chlorophyll-derived dietary phototoxin and protoporphyria. *Proc. Natl Acad. Sci. USA* **99,** 15649–15654 (2002).
45. Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4,** 1–7 (1994).
46. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404,** 193–197 (2000).
47. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173,** 1213–1225 (1991).



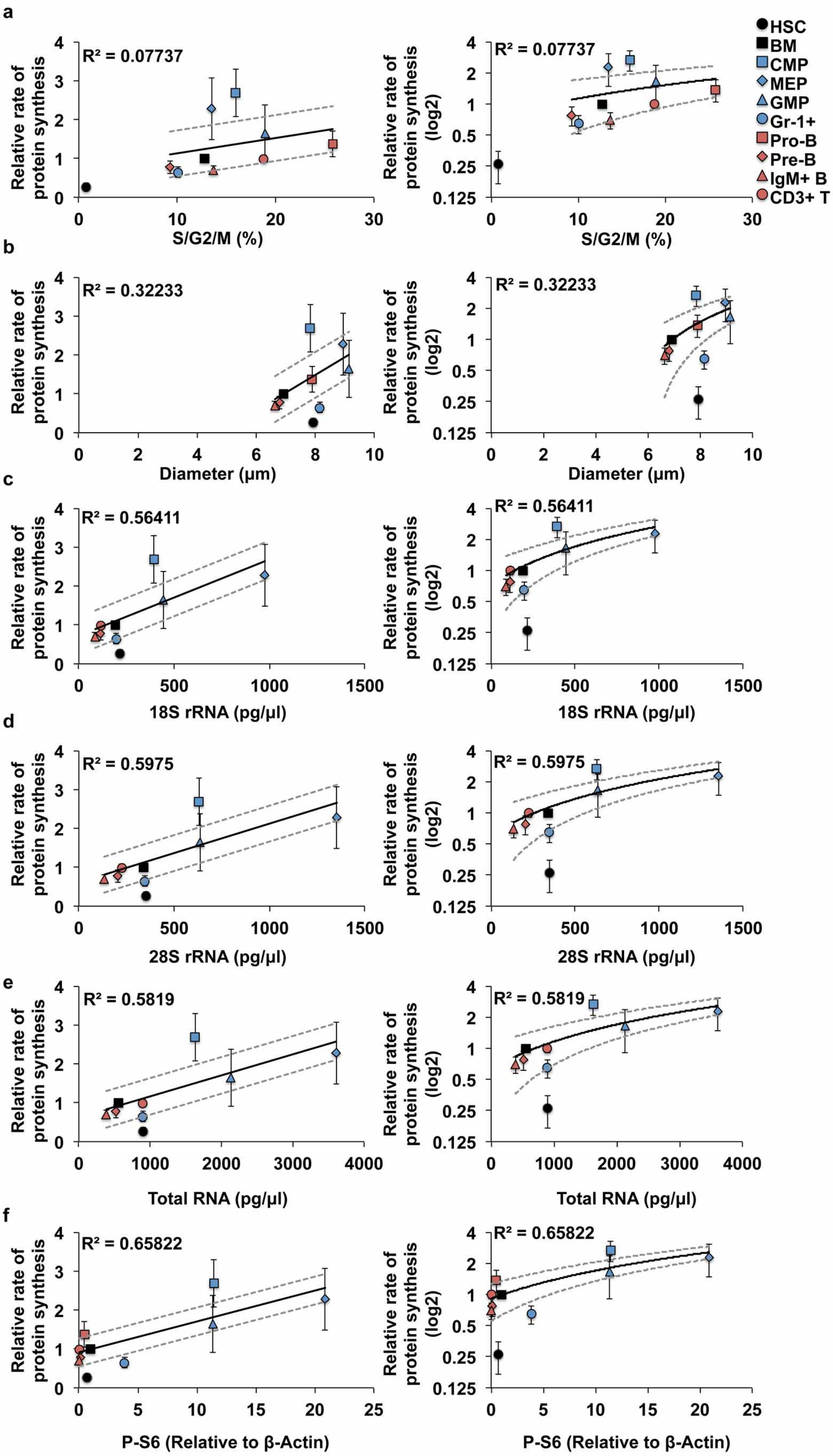
Extended Data Figure 1 | Isolation of haematopoietic progenitor cell populations by flow cytometry and histograms showing protein synthesis *in vivo* relative to HSCs from the same mice. **a**–**e**, One hour after OP-Puro administration to mice we observed no effect on bone marrow cellularity (one femur and one tibia (**a**); $n = 7$ PBS treated and $n = 9$ OP-Puro treated mice) or the frequencies of CD150$^+$CD48$^-$LSK HSCs (**b**; $n = 4$ PBS treated, $n = 6$ OP-Puro treated mice), annexin V$^+$ bone marrow cells (**c**; $n = 4$ PBS treated mice, $n = 6$ OP-Puro treated mice), annexin V$^+$ HSCs (**d**; $n = 4$ PBS treated mice, $n = 6$ OP-Puro treated mice), or HSCs in S/G$_2$/M phase of the cell cycle (**e**; $n = 3$ mice per treatment; **a**–**e** each reflect two or three independent experiments). **f**–**i**, Representative flow-cytometry plots showing the markers and gating strategies used to isolate CMPs[46], GMPs[46], and MEPs[46] (**f**), pro-B[47], pre-B[47] and IgM$^+$ B cells (**g**), Gr-1$^+$ myeloid cells (**h**), CD3$^+$ T cells (**h**) and CD71$^+$Ter119$^+$ erythroid progenitors (**i**). Each panel also shows OP-Puro incorporation histograms for each cell population relative to HSCs after 1 h of OP-Puro incorporation *in vivo*. The level of background fluorescence from PBS treated controls is overlaid in black. **j**, Data from Fig. 1h showing protein synthesis in various haematopoietic-cell populations relative to unfractionated bone marrow cells on a log$_2$ scale ($n = 15$ mice from 9 independent experiments). All data represent mean ± s.d. Two-tailed Student's t-tests were used to assess statistical significance in **a**–**e**. The statistical significance of differences relative to HSCs in **j** was assessed using a repeated-measures one-way ANOVA followed by Dunnett's test for multiple comparisons. Asterisks indicate statistical comparison to HSCs (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).



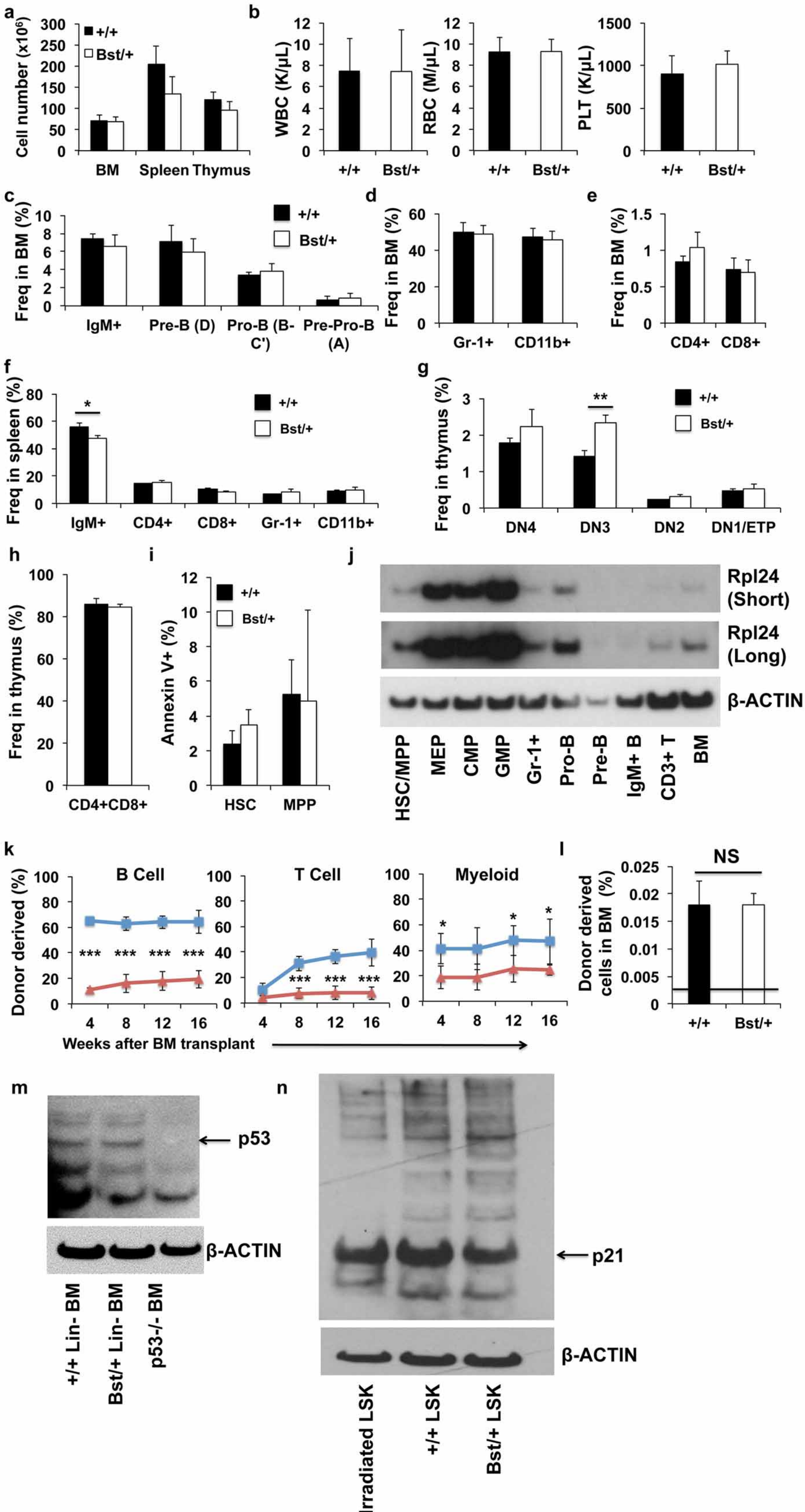
Extended Data Figure 2 | OP-Puro-containing polypeptides are not degraded within 30 min, the degradation that occurs over 24 h is blocked by bortezomib, and OP-Puro administration does not induce cell death. **a**, OP-Puro fluorescence in haematopoietic cells after 1 h of OP-Puro administration *in vivo* followed by a 30-min *ex vivo* incubation on ice or at 37 °C ($n = 11$ mice from 4 independent experiments). **b**, OP-Puro fluorescence in haematopoietic cells 24 h after OP-Puro administration *in vivo*. Treatment with bortezomib 1 h before OP-Puro administration increased OP-Puro fluorescence in every cell population 24 h later ($n = 3$ independent experiments; total number of mice per treatment are shown in the panel). **c**, Frequency of annexin V$^+$ cells in each cell population 1 h after OP-Puro administration *in vivo* relative to the same cells from untreated mice ($n = 7$ mice per treatment from 2 independent experiments). All data represent mean ± s.d. To assess the statistical significance of treatment effects within the same cell population we performed two-tailed Student's *t*-tests (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). To assess the statistical significance of differences between HSCs and other cell populations in **a**, we performed a repeated-measures one-way ANOVA followed by Dunnett's test for multiple comparisons (†$P < 0.05$, ††$P < 0.01$, †††$P < 0.001$).



Extended Data Figure 3 | Cyclophosphamide and GCSF treatment drives certain cells into cycle and increases protein synthesis. **a**, Frequency of dividing cells in S/G_2/M phases of the cell cycle (>2N (>diploid) DNA content; $n = 5$ mice from 3 independent experiments). **b**, Total protein isolated from 50,000 unfractionated bone marrow cells or Gr-1$^+$ cells in G_0/G_1 or S/G_2/M measured by BCA assay ($n = 3$). **c**, Total RNA content in 15,000 cells from each stem- or progenitor-cell population ($n = 3$ mice). **d**, The frequency of cycling (KI-67$^+$) HSCs increased dramatically after treatment with cyclophosphamide (Cy) and GCSF ($n = 5$ untreated mice and $n = 6$ mice treated with cyclophosphamide and GCSF, from 2 independent experiments, $P < 0.001$). **e**, Frequency of KI-67$^+$ cells in haematopoietic-cell populations before and after treatment with cyclophosphamide and GCSF ($n = 5$ untreated mice and $n = 6$ mice treated with cyclophosphamide and GCSF for BM, HSC and MPP, $n = 3$ mice per treatment for other cell populations). **f**, **g**, Protein synthesis in G_0/G_1 and S/G_2/M cells from untreated mice or mice treated with cyclophosphamide followed by two days of GCSF ($n = 10$ mice per treatment from 6 independent experiments). These data are the same as shown in Fig. 3b, d, e, shown together in this panel for comparison. The data are plotted on a linear scale in **f** and on a $\log_2$ scale in **g**. All data represent mean ± s.d. To assess the statistical significance of treatment effects within the same cell population (**b**, **e**–**g**) we performed two-tailed Student's *t*-tests (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). To assess the statistical significance of differences between HSCs and each other cell population (**a**, **c**, **f**, **g**) we performed a repeated-measures one-way ANOVA followed by Dunnett's test for multiple comparisons (†$P < 0.05$, ††$P < 0.01$, †††$P < 0.001$).

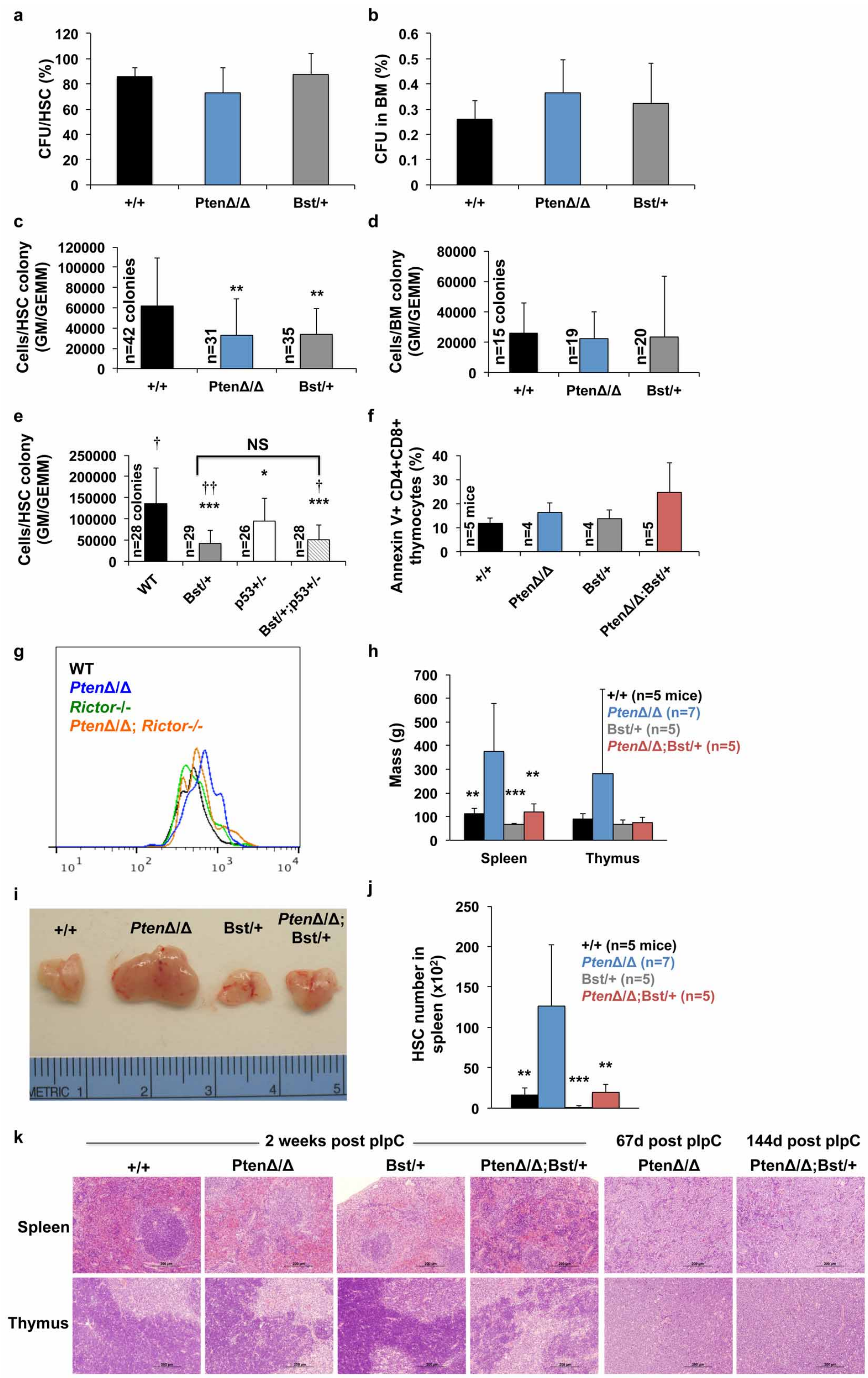


Extended Data Figure 4 | Differences in protein synthesis among haematopoietic stem and progenitor cells are not fully explained by differences in cell division, cell diameter, pS6 levels, rRNA or total RNA content. **a**–**f**, Scatter plots show the relative rates of protein synthesis (per hour) in each cell population (from Fig. 1h) plotted against the frequency of dividing cells (**a**; from Extended Data Fig. 3a), cell diameter (**b**; from Fig. 3f), 18S rRNA content (**c**; from Fig. 3g), 28S rRNA content (**d**; from Fig. 3g), total RNA content (**e**; from Extended Data Fig. 3c) and pS6 levels (**f**; from Fig. 5a normalized to β-actin). For each parameter, regressions were performed using all populations excluding HSCs and 95% confidence intervals were determined. R^2 values are shown in each plot. Rates of protein synthesis are plotted on a linear scale (left panels) and on a $\log_2$ scale (right panels). Note that HSCs were outliers with respect to each regression. $CD150^+CD48^-$LSK cells were used to determine HSC rates of protein synthesis, cell diameter and percentage $S/G_2/M$, and $CD48^-$LSK cells (HSCs and MPPs) were used to determine 18S, 28S, total RNA and pS6 levels (as these measurements required more cells). All data represent mean ± s.d.

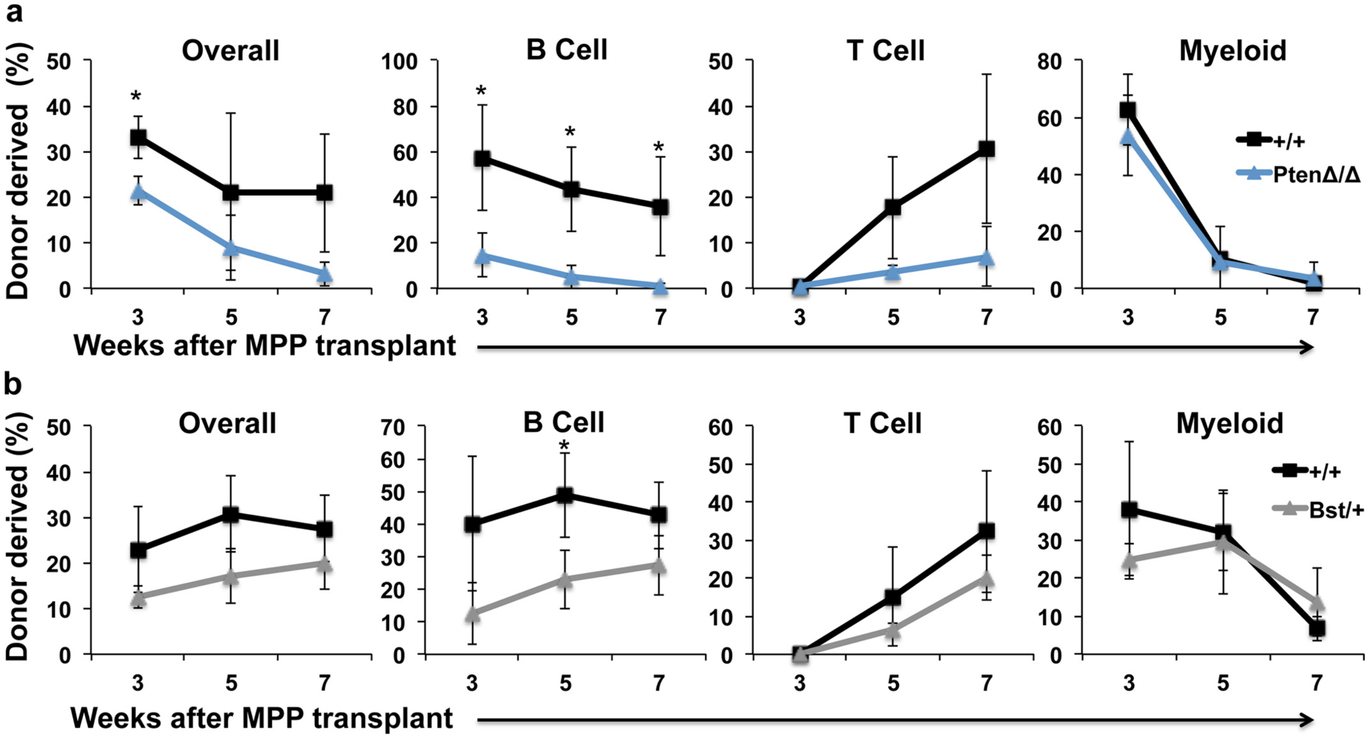


a
Cell number (x10^6)
+/+
Bst/+
BM
Spleen
Thymus
b
WBC (K/μL)
RBC (M/μL)
PLT (K/μL)
c
Freq in BM (%)
IgM+
Pre-B (D)
Pro-B (B-C')
Pre-Pro-B (A)
d
Gr-1+
CD11b+
e
CD4+
CD8+
f
Freq in spleen (%)
g
Freq in thymus (%)
DN4
DN3
DN2
DN1/ETP
h
CD4+CD8+
i
Annexin V+ (%)
HSC
MPP
j
Rpl24 (Short)
Rpl24 (Long)
β-ACTIN
HSC/MPP
MEP
CMP
GMP
Gr-1+
Pro-B
Pre-B
IgM+ B
CD3+ T
BM
k
Donor derived (%)
B Cell
T Cell
Myeloid
Weeks after BM transplant
l
NS
Donor derived cells in BM (%)
m
p53
+/+ Lin- BM
Bst/+ Lin- BM
p53-/- BM
n
p21
Irradiated LSK
+/+ LSK
Bst/+ LSK

Extended Data Figure 5 | *Rpl24*$^{Bst/+}$ mice have normal frequencies of lymphoid and myeloid lineage progenitors and do not express increased p53 or p21^{Cip1} in adult haematopoietic cells. **a**, Bone marrow (2 femurs and 2 tibias; $n = 5$ wild-type and $n = 6$ *Rpl24*$^{Bst/+}$ mice from 4 experiments), spleen ($n = 3$ wild-type and $n = 4$ *Rpl24*$^{Bst/+}$ mice from 2 experiments), and thymus cellularity ($n = 3$ wild-type and $n = 4$ *Rpl24*$^{Bst/+}$ mice from 2 experiments). **b**, White blood cell, red blood cell and platelet counts ($n = 5$ wild-type and $n = 6$ *Rpl24*$^{Bst/+}$ mice from 4 experiments). **c–f**, The frequencies of B (**c**), myeloid (**d**) and T (**e**) lineage cells in the bone marrow and spleen (**f**) of *Rpl24*$^{Bst/+}$ and control mice ($n = 3$ wild-type and $n = 4$ *Rpl24*$^{Bst/+}$ mice from 2 experiments). **g**, **h**, The frequencies of T-lineage progenitors in the thymus of *Rpl24*$^{Bst/+}$ and control mice ($n = 3$ wild-type and $n = 4$ *Rpl24*$^{Bst/+}$ mice from 2 experiments). Double negative (DN)1 and early T-lineage progenitor (ETP) cells were $CD4^-CD8^-CD44^+CD25^-$; DN2 cells were $CD4^-CD8^-CD44^+CD25^+$; DN3 cells were $CD4^-CD8^-CD44^-CD25^+$; and DN4 cells were $CD4^-CD8^-CD44^-CD25^-$. **i**, The frequencies of annexin V^+ HSCs and MPPs in *Rpl24*$^{Bst/+}$ versus littermate control mice ($n = 3$ wild-type and $n = 4$ *Rpl24*$^{Bst/+}$ mice from 3 experiments). **j**, Western blot analysis for Rpl24 and β-actin using 30,000 cells from each haematopoietic-cell population. Differences in β-actin between lanes represent differences in β-actin content per cell (one representative blot from two independent experiments). **k**, Donor- B-cell, T-cell and myeloid-cell engraftment when 5×10^5 donor bone marrow cells were transplanted along with 5×10^5 recipient bone marrow cells into irradiated recipient mice ($n = 4$ independent experiments with a total of 17 recipients for wild-type cells and 20 for *Rpl24*$^{Bst/+}$ cells). These transplant recipients are the same as those shown in Fig. 4e. **l**, The frequency of donor cells in the bone marrow 20 h after transplanting 1×10^5 donor LSK cells from *Rpl24*$^{Bst/+}$ or wild-type control mice into irradiated recipient mice ($n = 3$ recipients per donor). The horizontal line represents the level of background detected in an untransplanted control. **m**, Western blot analysis for p53 using 5×10^5 Lineage$^-$ bone marrow cells from wild-type or *Rpl24*$^{Bst/+}$ mice, or 5×10^5 bone marrow cells from a *p53*$^{-/-}$ mouse (one representative blot from two experiments). **n**, Western blot analysis for p21^{Cip1} using 285,000 LSK cells from the bone marrow of adult wild-type or *Rpl24*$^{Bst/+}$ mice, or 142,500 LSK cells from a wild-type mouse that received 540 rad of total body irradiation 3 to 4 h before being euthanized (one representative blot from three independent experiments). All data represent mean ± s.d. Two-tailed Student's *t*-tests were used to assess statistical significance (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).



Extended Data Figure 6 | *Rpl24*$^{Bst/+}$ and *Pten*-deficient progenitors form colonies with normal cellularity but *Rpl24*$^{Bst/+}$ impairs the development of haematopoietic neoplasms after *Pten* deletion. **a**, **b**, The percentage of HSCs (**a**) or bone marrow cells (**b**) that formed colonies in methylcellulose within 14 days of culture ($n = 3$ mice per genotype in 3 independent experiments with 16 HSCs or 3,200 bone marrow cells tested per mouse per experiment). **c**, **d**, The average number of cells per granulocyte-monocyte (GM) or granulocyte, erythrocyte, monocyte, megakaryocyte (GEMM) colony derived from single HSCs (**c**) or bone marrow cells plated at clonal density (**d**) ($n = 4$ independent experiments). **e**, The average number of cells per GM or GEMM colony derived from individual HSCs of the indicated genotypes 15 days after plating ($n = 2$ independent experiments). **f**, Frequency of annexin V^{+} CD4^{+}CD8^{+} thymocytes ($n = 5$ independent experiments). **g**, Representative histograms of OP-Puro fluorescence in HSCs of the indicated genotypes. **h**, Mass of spleens and thymuses 2 weeks after pIpC administration ($n = 7$ independent experiments). **i**, Representative photographs of thymuses 2 weeks after pIpC administration to wild-type, *Mx1-Cre; Pten*$^{fl/fl}$, *Rpl24*$^{Bst/+}$ and *Mx1-Cre; Pten*$^{fl/fl}$; *Rpl24*$^{Bst/+}$ mice. **j**, HSCs in the spleen 2 weeks after pIpC administration ($n = 7$ independent experiments). **k**, Haematoxylin and eosin stained spleen and thymus sections from mice 2 weeks after pIpC administration or when they were killed owing to illness. All data represent mean ± s.d. In **a**–**d** and **f**, two-tailed Student's *t*-tests were used to assess statistical significance relative to wild-type; $*P < 0.05$, $**P < 0.01$. To assess statistical significance in **e** we performed a one-way ANOVA followed by Tukey's *t*-tests for multiple comparisons (relative to wild-type, $*P < 0.05$; $***P < 0.001$; and relative to $p53^{+/-}$, $\dagger P < 0.05$, $\dagger\dagger P < 0.01$). To compare the statistical significance of differences among genotypes in **h** and **j** we performed a one-way ANOVA followed by Dunnett's test for multiple comparisons relative to *Pten*-deficient ($**P < 0.01$, $***P < 0.001$).



Extended Data Figure 7 | *Rpl24*$^{Bst/+}$ and *Pten*-deficient MPPs have relatively normal reconstituting activity. **a**, **b**, 100 donor CD150^{-}CD48^{-}LSK MPPs from *Mx1-Cre; Pten*$^{fl/fl}$ versus control mice (**a**; $n = 3$ independent experiments with a total of 13 recipients per genotype) or *Rpl24*$^{Bst/+}$ versus control mice (**b**; $n = 3$ independent experiments with a total of 14 recipients of wild-type cells and 13 recipients of *Rpl24*$^{Bst/+}$ cells) were transplanted along with 3×10^5 recipient-type bone marrow cells into irradiated recipient mice. Donor-cell engraftment levels in the peripheral blood were assessed at 3, 5 and 7 weeks after transplantation. All data represent mean ± s.d. Two-tailed Student's *t*-tests were used to assess statistical significance relative to wild-type; $^*P < 0.05$.

doi:10.1038/nature13036

Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies

Nicole A. Doria-Rose[1]*, Chaim A. Schramm[2]*, Jason Gorman[1]*, Penny L. Moore[3,4,5]*, Jinal N. Bhiman[3,4], Brandon J. DeKosky[6], Michael J. Ernandes[1], Ivelin S. Georgiev[1], Helen J. Kim[7,8,9], Marie Pancera[1], Ryan P. Staupe[1], Han R. Altae-Tran[1], Robert T. Bailer[1], Ema T. Crooks[10], Albert Cupo[11], Aliaksandr Druz[1], Nigel J. Garrett[5], Kam H. Hoi[12], Rui Kong[1], Mark K. Louder[1], Nancy S. Longo[1], Krisha McKee[1], Molati Nonyane[3], Sijy O'Dell[1], Ryan S. Roark[1], Rebecca S. Rudicell[1], Stephen D. Schmidt[1], Daniel J. Sheward[13], Cinque Soto[1], Constantinos Kurt Wibmer[3,4], Yongping Yang[1], Zhenhai Zhang[2], NISC Comparative Sequencing Program†, James C. Mullikin[14,15], James M. Binley[10], Rogier W. Sanders[16], Ian A. Wilson[7,8,9,17], John P. Moore[11], Andrew B. Ward[7,8,9], George Georgiou[6,12,18], Carolyn Williamson[5,13], Salim S. Abdool Karim[5,19], Lynn Morris[3,4,5], Peter D. Kwong[1], Lawrence Shapiro[1,2] & John R. Mascola[1]

Antibodies capable of neutralizing HIV-1 often target variable regions 1 and 2 (V1V2) of the HIV-1 envelope, but the mechanism of their elicitation has been unclear. Here we define the developmental pathway by which such antibodies are generated and acquire the requisite molecular characteristics for neutralization. Twelve somatically related neutralizing antibodies (CAP256-VRC26.01–12) were isolated from donor CAP256 (from the Centre for the AIDS Programme of Research in South Africa (CAPRISA)); each antibody contained the protruding tyrosine-sulphated, anionic antigen-binding loop (complementarity-determining region (CDR) H3) characteristic of this category of antibodies. Their unmutated ancestor emerged between weeks 30–38 post-infection with a 35-residue CDR H3, and neutralized the virus that superinfected this individual 15 weeks after initial infection. Improved neutralization breadth and potency occurred by week 59 with modest affinity maturation, and was preceded by extensive diversification of the virus population. HIV-1 V1V2-directed neutralizing antibodies can thus develop relatively rapidly through initial selection of B cells with a long CDR H3, and limited subsequent somatic hypermutation. These data provide important insights relevant to HIV-1 vaccine development.

Developmental pathways of antibodies that neutralize HIV-1 represent potential templates to guide vaccine strategies, if their constituent molecular events were understood and could be reproduced[1–3]. Almost all HIV-1 infected individuals mount a potent antibody response within months of infection, but this response preferentially neutralizes autologous virus, which rapidly escapes[4,5]. Cross-reactive antibodies capable of neutralizing most HIV-1 strains arise in only ~20% of donors after 2–3 years of infection[6–9]. An understanding of the development of broadly neutralizing antibody (NAb) lineages in such donors could provide a roadmap for vaccine design.

One means to obtain such a roadmap is through isolation of broadly cross-reactive neutralizing antibodies, characterization of their genetic sequence and molecular properties, and examination of the B cell genetic record with next-generation sequencing (NGS)[10–14]. The greatest insights can be gained with longitudinal sampling from early after the time of HIV-1 infection[15]. This allows for a genetic delineation of the molecular evolution leading from an unmutated ancestor antibody, through affinity maturation, to acquisition of neutralization breadth. In principle, such a roadmap should link antibody molecular characteristics to the genetic development that a successful vaccine would retrace.

Neutralizing antibodies to the V1V2 region of the HIV-1 viral spike are among the most prevalent cross-reactive antibodies elicited by natural infection[6,16–18] and have been isolated from several donors[19–21]. These antibodies have long heavy-chain complementarity-determining region 3 loops (CDR H3s) that are protruding, anionic and often tyrosine sulphated[22,23]. These CDR H3s penetrate the HIV-1 glycan shield, recognizing a quaternary glycopeptide epitope at the apex of the HIV-1 spike that is formed by V1V2s from at least two gp120 protomers[22–24]. Here we use antibody isolation, B-cell next-generation sequencing, structural characterization, and viral single-genome amplification (SGA) to delineate longitudinal interactions between the developing antibody and autologous virus within donor CAP256, who showed evidence of V1V2-mediated neutralization breadth after one year[18,25,26]. Our results define the molecular requirements and genetic pathways that lead to V1V2-directed neutralization, providing a template for their vaccine elicitation.

[1]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]Department of Biochemistry, Columbia University, New York, New York 10032, USA. [3]Center for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service (NHLS), Johannesburg, 2131, South Africa. [4]Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2050, South Africa. [5]Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Congella, 4013, South Africa. [6]Department of Chemical Engineering, University of Texas at Austin, Austin, Texas 78712, USA. [7]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, USA. [8]Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, California 92037, USA. [9]IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California 92037, USA. [10]Torrey Pines Institute, San Diego, California 92037, USA. [11]Weill Medical College of Cornell University, New York, New York 10065, USA. [12]Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas, USA. [13]Institute of Infectious Diseases and Molecular Medicine, Division of Medical Virology, University of Cape Town and NHLS, Cape Town 7701, South Africa. [14]NISC Comparative Sequencing program, National Institutes of Health, Bethesda, Maryland 20892, USA. [15]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [16]Department of Medical Microbiology, Academic Medical Center, Amsterdam 1105 AZ, Netherlands. [17]Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA. [18]Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas 78712, USA. [19]Department of Epidemiology, Columbia University, New York, New York 10032, USA.
*These authors contributed equally to this work.
†A list of authors and their affiliations appears in the Supplementary Information.

Antibody isolation and characterization

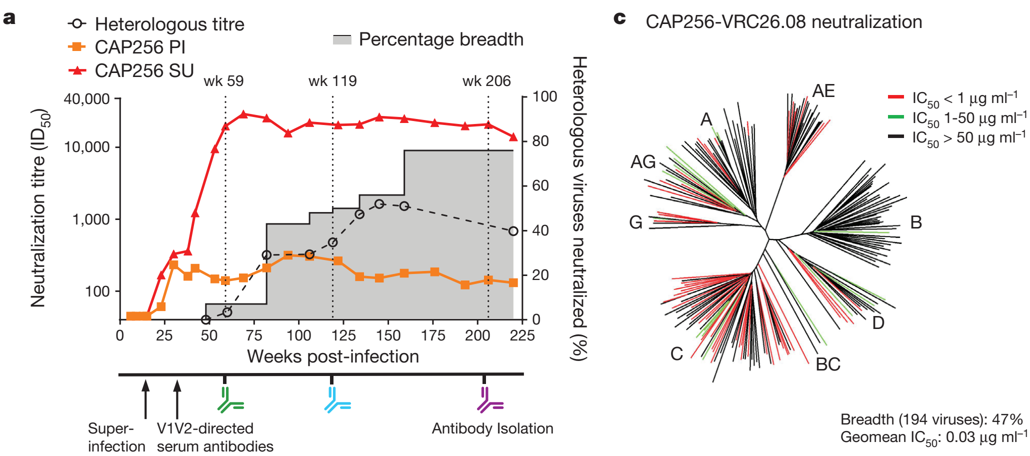
Donor CAP256 peripheral blood mononuclear cells (PBMCs) sampled 59, 119 and 206 weeks post-infection were used to isolate 12 monoclonal antibodies by high-throughput B-cell culture, functional screening by microneutralization, and PCR with reverse transcription (RT–PCR) of antibody variable regions[27,28] (Fig. 1a). All 12 were somatically related and distinguished by long CDR H3s of 35–37 amino acids (Kabat[29] numbering) (Fig. 1b and Extended Data Fig. 1a). The heavy and light chains exhibited somatic mutation of 4–15% from their germline-encoded *V*-genes, VH3-30 and Vλ1-51, respectively (Extended Data Fig 1 and Extended Data Table 1). When these antibodies were reconstituted as IgG1s, they showed varying degrees of heterologous virus neutralization and were extremely potent against many subtype A and C strains (Fig. 1b, c, Extended Data Fig. 2 and Supplementary Fig. 1). The combination of all 12 antibodies recapitulated plasma neutralization (Supplementary Fig. 2), indicating the CAP256-VRC26 antibody lineage to be responsible for the neutralization breadth and potency of donor CAP256.

To map the epitope of the CAP256-VRC26 antibodies, we used neutralization fingerprints[18]; binding assays for HIV-1 Envelope (Env) in soluble, cell surface[30], and viral particle[31] contexts; and negative stain electron microscopy (EM) of Fab CAP256-VRC26.09 bound to a soluble cleaved version of the HIV-1 trimer[24,32,33] (Fig. 2a–c, Supplementary Fig. 3 and Extended Data Figs 3 and 4). Recognition of Env by CAP256-VRC26 antibodies was similar to PG9-class neutralizing antibodies that recognize the trimeric V1V2 cap[24], with high specificity for the Env native quaternary conformation and one Fab bound per trimer (Fig. 2c, left and Extended Data Fig. 4). Neutralization activity of CAP256-VRC26 antibodies was reduced or knocked out by Env mutations in V1V2 strands B and C (Fig. 2d), much like the CAP256 plasma[25,26] and PG9-class neutralizing antibodies[22,23,34]; although unlike PG9, the CAP256-VRC26 antibodies were only partially and variably sensitive to loss of glycans at N160 and N156 (Fig. 2d and Extended Data Fig. 5). Overall, these data indicated the epitope to be at the membrane-distal apex of the HIV-1 spike close to the trimer axis (Fig. 2e), providing a structural explanation for the observed quaternary specificity.

Origin and development of the lineage

To obtain a genetic record of the CAP256-VRC26 antibody lineage, we analysed B cell-immunoglobulin transcripts at eight time points between 15 and 206 weeks post-infection by 454 pyrosequencing. Although no CAP256-VRC26 lineage-related transcripts were detected at 15 and 30 weeks, related heavy chain and light chain transcripts were found at all later time points (Fig. 3a). To track longitudinal prevalence, we used identity-divergence plots[12] of all heavy chain reads assigned to the same VH3-30 germline gene as the isolated antibodies. Using CAP256-VRC26.01 or CAP256-VRC26.08 as the identity referents, segregated islands of related heavy chain sequences first appeared at week 38 (Fig. 3b). For all 12 antibodies, the prevalence and identity of related sequences peaked close to the time of the antibody isolation (Supplementary Fig. 4). To obtain additional antibody lineage data, we performed linked VH:VL paired sequencing[35] at five time points (Fig. 3a and Supplementary Table 1). Of 157 unique CAP256-VRC26 pairs, 7 matched either heavy or light chain sequences present in the 454 pyrosequencing data, including 2 for which both heavy and light chain sequences had previously been captured (Fig. 3c).

Maximum-likelihood phylogenetic trees were constructed using the isolated antibodies and the 454 data (Fig. 3c). The lineage bifurcates early, with one branch leading to CAP256-VRC26.01 and a second developing into CAP256-VRC26.02–12. The unmutated common ancestors (UCAs) for the heavy and light chain were inferred from the phylogenetic trees (Fig. 3c). For the light chain, the UCA had a 12-residue CDR L3, as in CAP256-VRC26.01, and for the heavy chain, the inferred UCA had a 35-residue CDR H3 (Extended Data Fig. 6), probably the result of VDJ recombination with a single *D*-gene, IgHD3-3*01 and non-templated (N)-nucleotide insertions of 34 and 31 nucleotides at each



b

Week	Antibody	Somatic mutations (nt) Vλ1-51*02	Somatic mutations (nt) VH3-30*18	CDR H3 sequence	CDR H3 length (aa)	Neutralization (47 strains) Breadth	Neutralization (47 strains) Potency (IC_{50}, µg ml^{-1})
59	CAP256-VRC26.01	3.9%	8.3%	CAKDVGDYKSDEWGT-EYYDISISYPIQDPRAM--VGAFDLW	35	19%	1.88
	CAP256-VRC26.02	4.9%	8.7%	CAKDIREYECEYWTS-DYYDFGRPQPCIDSRGV--VGTFDVW	35	17%	0.40
	CAP256-VRC26.03	7.4%	8.7%	CAKDLREDECEEWWS-DYYDFGKQLPCRKSRGV--AGIFDGW	35	36%	0.08
	CAP256-VRC26.04	8.1%	9.0%	CAKDLREDECEEWWS-DYYDFGKQLPCRKSRGV--AGIFDKW	35	30%	0.32
119	CAP256-VRC26.05	5.4%	10.1%	CARDQRYYECEEWAS-DYYDFGREQPCLDPRGV--VGIFDLW	35	21%	0.10
	CAP256-VRC26.06	7.4%	10.8%	CARDLRELECEEWTLYNYYDFGSRGPCVDPRGV--AGSFDVW	36	17%	0.38
	CAP256-VRC26.07	7.7%	11.8%	CAKDLREDECEEWWS-DYYDFGKKLPCRKSRGV--AGVFDKW	35	13%	1.51
	CAP256-VRC26.08	9.8%	11.8%	CVRDQREDECEEWWS-DYYDFGRELPCRKFRGLGLAGIFDIW	37	47%	0.11
	CAP256-VRC26.09	9.8%	14.2%	CVKDQREDECEEWWS-DYYDFGRELPCRKSRGLGLAGIFDMW	37	47%	0.07
	CAP256-VRC26.10	3.9%	11.8%	CAKDMREYECEYWTS-DYYDFGRPQPCIDRRGV--VGIFDMW	35	23%	0.60
206	CAP256-VRC26.11	13.7%	11.9%	CVKDMRELECEEWAS-DYYDFGKPQPCLDRRGV--SGISAWW	35	26%	0.94
	CAP256-VRC26.12	8.4%	15.3%	CARDLRESECEEWES-DYYDFGKKGPCVKPRGV--AGGLDLW	35	6%	0.49

Figure 1 | Development of broad neutralization by donor CAP256 and isolation of neutralizing antibodies. **a**, Timing of antibody isolation in relation to plasma neutralization titres against the primary infecting virus (PI), the superinfecting virus (SU), and a panel of 40 heterologous viruses (geometric mean titre shown). Percentage breadth (grey area), percentage of viruses neutralized with plasma median inhibitory dilution (ID_{50}) > 45. **b**, Genetic characteristics and neutralization breadth and potency of the 12 isolated antibodies. Week of antibody isolation and *V*-gene mutation rates are indicated. Residues flanking the Kabat-defined CDR H3 sequences are shown in light grey. Neutralization was assessed against a panel of 47 heterologous viruses. **c**, Breadth and potency of antibody CAP256-VRC26.08 on a panel of 194 Env-pseudoviruses. Dendrogram shows phylogenetic relatedness of Env sequences in the panel.

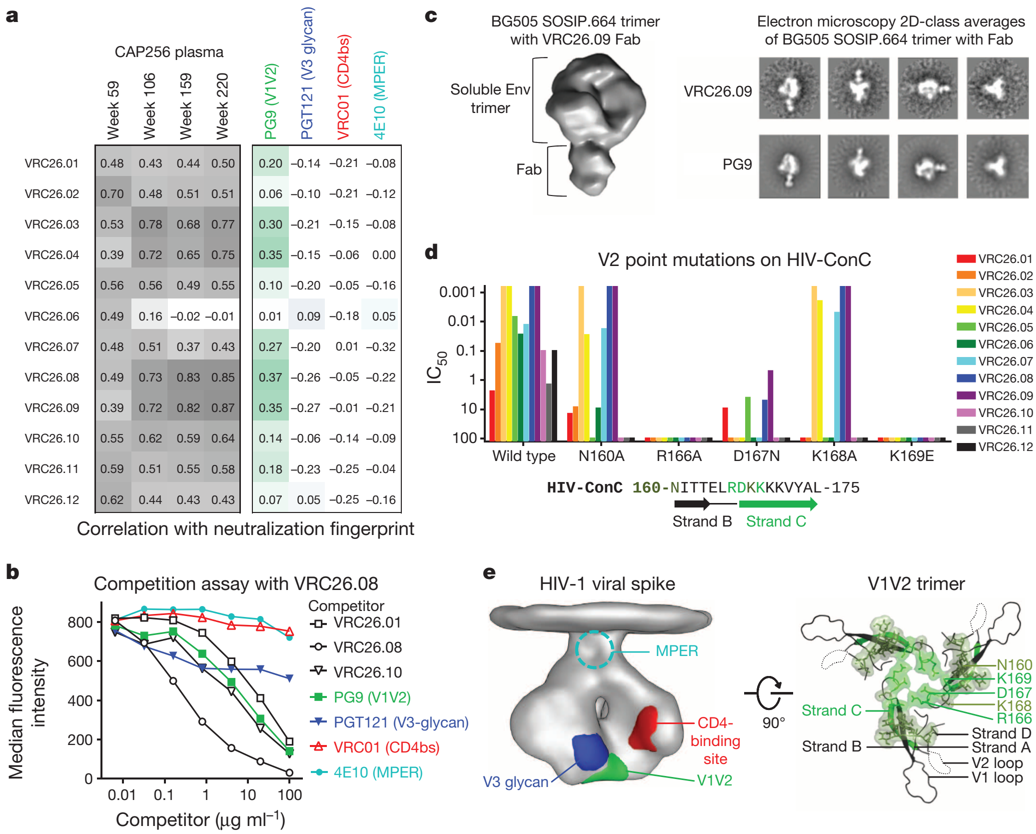


Figure 2 | Mapping of CAP256-VRC26 epitope on the HIV-1 Env spike. **a**, Correlations between neutralization fingerprints (see Methods) of CAP256-VRC26 antibodies and CAP256 plasma (left). Darker grey indicates stronger correlation. Correlations between neutralization fingerprints of CAP256-VRC26 antibodies and representative antibodies targeting the major HIV-1 neutralization epitopes (right). Correlations are colour-coded by antibody; darker shades indicate stronger correlations. **b**, Competition assay. Binding to ZM53-Env-expressing 293T cells by labelled CAP256-VRC26.08 and unlabelled competitor antibodies measured by flow cytometry. Assay shown is representative of three experiments. **c**, Negative stain electron microscopy (EM) 3D reconstruction of CAP256-VRC26.09 Fab in complex with soluble cleaved BG505 SOSIP.664 trimer (left); 2D-class averages of VRC26.09 and PG9 in complex with BG505 SOSIP.664 trimer (right). **d**, Neutralization of Env-pseudoviruses with HIV-ConC and V2 point mutants. Sequence shows amino acids 160–175. **e**, Location of HIV-1 epitopes. EM density of viral spike[50], with viral membrane at top and major sites of vulnerability shown as determined by structural mapping of antibody interactions[24] (left). The gp41 membrane proximal external region (MPER) is shown schematically. Model of V1V2 based on EM structure of BG505 SOSIP.664 trimer[24,32], viewed looking towards the viral membrane along the trimer axis (right). Green ribbon, strand C. V2 mutations from panel **d** are shown with surface representation; brighter green indicates more potent effects on neutralization.

junction (Supplementary Fig. 5). This inferred UCA was further supported by very-low-divergence sequences among the lineage members identified from the week 38 heavy chain data. Five unique sequences were found, all of which had CDR H3s matching the inferred UCA in at least 30 of 35 amino acids while containing three or fewer nucleotide changes in VH and JH combined (Extended Data Fig. 6). Thus, the longitudinal NGS analysis established the first appearance of the CAP256-VRC26 lineage; defined the UCA, the product of gene recombination in the ancestor B cell of the lineage; and provided a genetic record of the development of this lineage over four years.

Structures of CAP256-VRC26 antibodies

To define the structural characteristics of CAP256-VRC26 lineage development, we determined crystal structures for Fabs of the UCA and six antibodies from weeks 59, 119 and 206 (Fig. 4, Supplementary Table 2, and Supplementary Fig. 6a). The mature CDR H3s protruded ~20Å above the antigen-combining surface of the heavy chain and contained a 2-stranded β-sheet, *O*-sulphated tyrosines, and an intra-CDR H3 disulphide bond (Fig. 4a, b). The CDR H3s of the UCA and CAP256-VRC26.01 lacked a CDR H3 disulphide bond, exhibited greater disorder and were positioned more proximal to the light chain (Fig. 4c); the appearance of the disulphide bond correlated with adoption of the mature CDR H3 orientation (Fig. 4c, Supplementary Fig. 6b, and Extended Data Fig. 7a). Mutation to remove the relevant cysteine residues in VRC26.03 resulted in loss of neutralization potency and breadth (Extended Data Fig. 7b, c). Additionally, the appearance of CDR H3 cysteines coincided with a glycine to arginine mutation at the base of the CDR H3, possibly limiting flexibility of the mature antibodies (Extended Data Fig. 7a, b and Supplementary Fig. 7). Overall, the CAP256-VRC26 lineage begins with an anionic protruding CDR H3 with structural properties similar to previously determined V1V2-directed broadly neutralizing antibodies. Development over four years involves the introduction of almost 20 light chain and over 30 heavy chain mutations, including a disulphide bond. The CDR H3 changes its overall orientation while losing negative charge and maintaining tyrosine sulphation (Fig. 4b, c, right).

HIV Env evolution during NAb development

To gain insight into the temporal HIV-1 Env changes driving the development of the CAP256-VRC26 lineage, we used SGA to determine viral sequences over ~3 years. CAP256 Env sequences showed high levels of diversity driven, in part, by recombination between the superinfecting virus (SU) that was first detected 15 weeks post-infection and the primary infecting virus (PI)[26] (Fig. 5a, Supplementary Figs 8, 9). Differences between the primary infecting virus and superinfecting virus Env sequences included V2 residues 165 and 169, and an N160 glycan in the superinfecting virus that was not present in the primary infecting virus (Fig. 5b and Extended Data Fig. 8a, b). Notably, compared to the primary

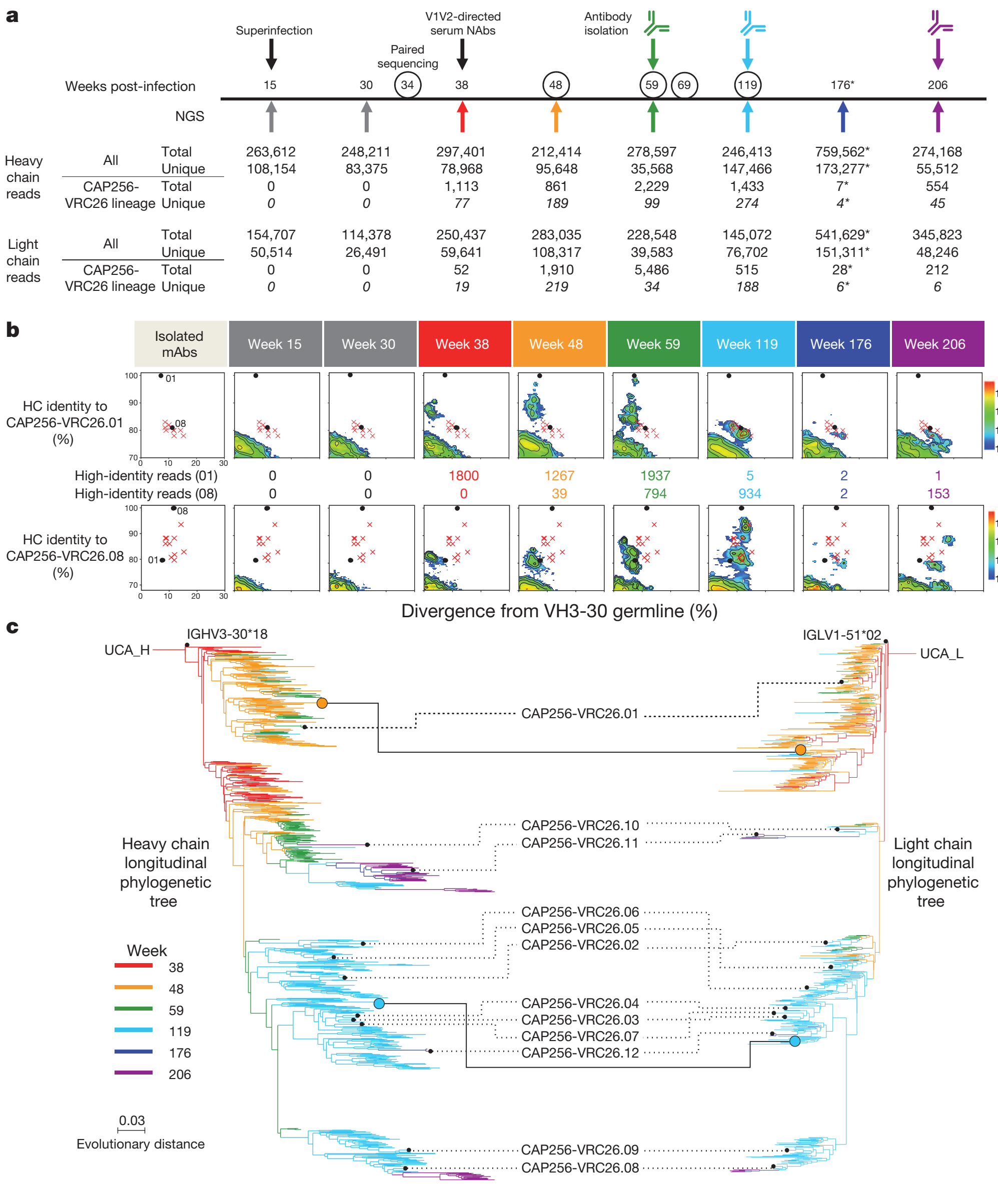


Figure 3 | Maturation of the CAP256-VRC26 lineage revealed by NGS and VH:VL paired sequencing of B cell transcripts. **a**, Timeline of longitudinal peripheral blood samples with quantification of all NGS sequence reads (total and unique), and CAP256-VRC26 lineage-related reads (total and unique). Arrows below the line indicate time points of 454 pyrosequencing for heavy and light chain sequences. Circles indicate time points of paired sequencing of sorted B cells (see Methods). PCR amplifications for pyrosequencing used primers specific for VH3 family sequences (heavy chain) and V lambda sequences (light chain), with the exception of the week 176 sample (asterisk), which was amplified using all-*VH* gene primers, resulting in fewer CAP256-VRC26 specific reads. **b**, Maturation time course for CAP256-VRC26.01 (top) and CAP256-VRC26.08 (bottom panels). Heat map plots show sequence identity (vertical axis) versus germline divergence (horizontal axis) for NGS data. The 12 isolated antibodies are displayed as red 'x' marks for reference, with the exception of the CAP256-VRC26.01 and 08 antibodies which are shown as black dots. Numbers between the top and bottom panels correspond to the number of raw reads with at least 85% identity to the indicated antibody (VRC26.01 (top), VRC26.08 (bottom)). **c**, Phylogenetic trees of the CAP256-VRC26 clonal lineage for heavy chain (left) and light chain (right) were constructed by maximum likelihood using the 454 sequences and the isolated antibodies (black dots, labelled with antibody name). Branches are coloured by time point when NGS sequences were first detected. The orange and blue circles indicate linked heavy and light chain sequences from the paired sequencing data. Scale, rate of nucleotide change (per site) between nodes.

infecting virus, the superinfecting virus contained V2 residues that are more commonly found among circulating viruses (Extended Data Fig. 8a). All 12 antibodies neutralized the superinfecting virus, and, with the exception of CAP256-VRC26.06, failed to neutralize the primary infecting virus, suggesting the superinfecting virus V1V2 initially engaged the naive B cell of the CAP256-VRC26 lineage (Fig. 5d, Extended Data Fig. 8c and Supplementary Fig. 10).

Before the CAP256-VRC26 antibodies developed, most Env sequences had V1V2 regions derived from the primary infecting virus (Fig. 5a–c and Supplementary Figs 8 and 9) and were therefore largely neutralization resistant (Fig. 5d and Supplementary Fig. 10). Among superinfecting-virus-like sequences, a rare K169I mutation arose under strong directional selection (Supplementary Table 3) as the CAP256-VRC26 lineage emerged, which rendered the superinfecting virus resistant to only the earliest antibody (Extended Data Fig. 8d, e), indicating that CAP256-VRC26.01-like antibodies drove this viral escape, followed by maturation of the lineage to tolerate I169. At 48 weeks, the viral population underwent a substantial shift (Fig. 5a and Supplementary Figs 8 and 9), with the

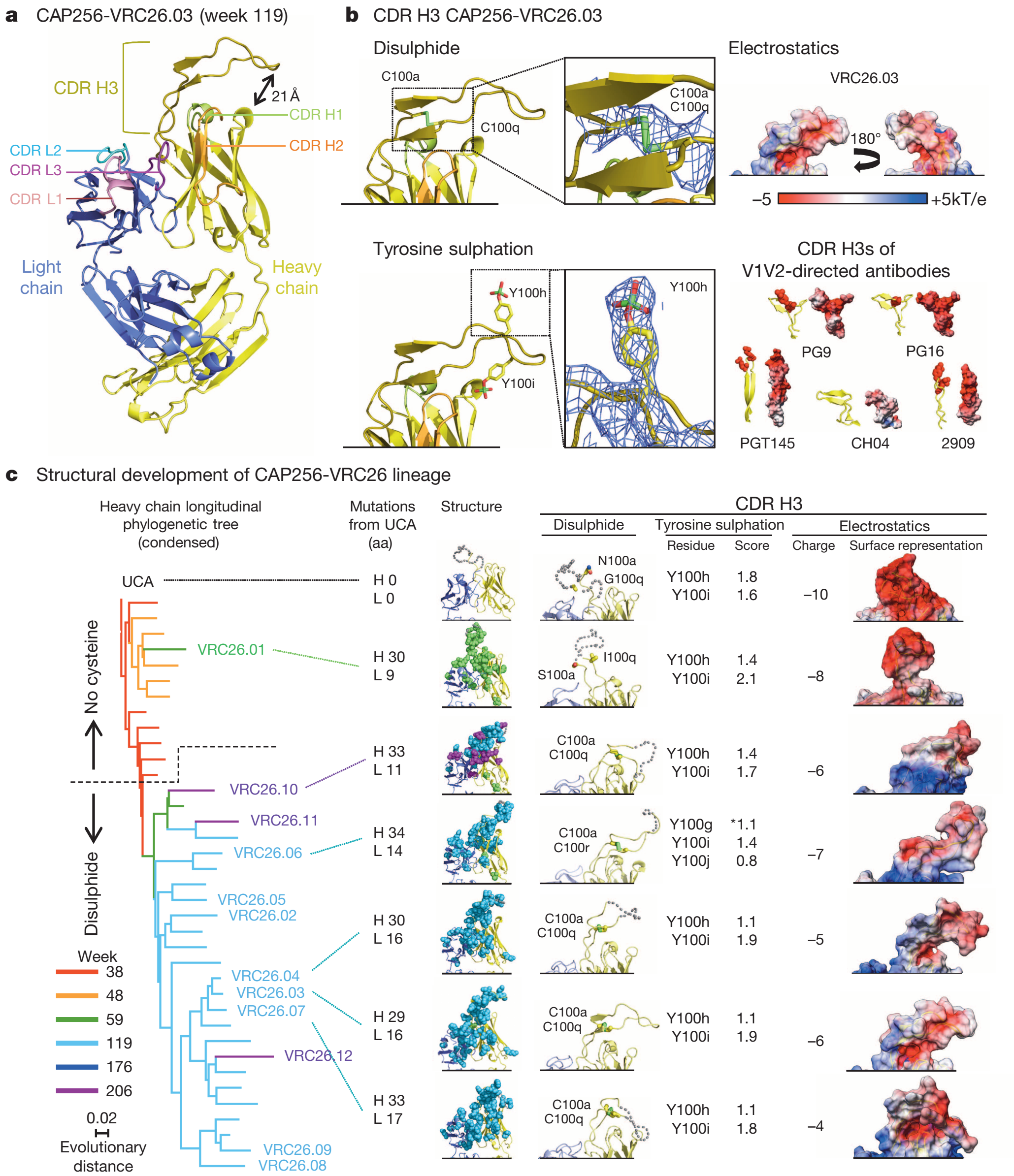


Figure 4 | Structural characteristics of the developing CAP256-VRC26 lineage. **a**, Crystal structure of the antigen-binding fragment (Fab) of CAP256-VRC26.03 shown in ribbon diagram representation. **b**, The intra-loop disulphide bond and tyrosine sulphation are shown in stick representation, and enlarged to show electron density (blue mesh, $2F_o$-F_c at 1σ) (left). Molecular surface, with electrostatic potentials coloured red for acidic and blue for basic (right). CDR H3 regions of broadly neutralizing V1V2-directed antibodies are shown for comparison, with the left image in ribbon representation (tyrosine sulphates highlighted in red) and the right image in electrostatic representation. **c**, A condensed heavy chain phylogenetic tree highlights the isolated antibodies (left). Scale, rate of nucleotide change between nodes. The number of amino acid (aa) mutations to the heavy chain (H) and light chain (L) relative to the UCA are shown. Structures of the variable regions (middle). Mutations from the UCA are represented as spheres coloured according to the week of antibody isolation at which the mutations first appear. CDR H3 details (right). Residues that are (or evolve to become) cysteines are labelled (grey dotted lines indicate modelled disordered regions). The position of tyrosines predicted to be sulphated (scores >1) are noted and were included in the formal charges shown for each CDR H3 and the electrostatic representations (far right). Asterisk denotes Tyr insertion in VRC26.06.

superinfecting-virus-like V1V2 dominating just before the development of neutralization breadth. Neutralization of Env clones by later antibodies (CAP256-VRC26.02-12) tracked with the presence of superinfecting-virus-like V1V2 sequences (black bar, Fig. 5c) until escape occurred through mutations at positions 166 or 169 (Fig. 5c, d and Extended Data Fig. 8d). These mutations resulted in a net charge change in the V2 epitope (+3 to 0, Fig. 5c, Extended Data Fig. 8b) concomitant with the antibody CDR H3s becoming less acidic over time (−10 to −4, Fig. 4 and Extended Data Fig. 9) suggesting co-evolution of the viral epitope and the antibody paratope. Overall, these results highlight the interplay between virus and antibody, with the superinfecting-virus-like V1V2 epitope stimulating expansion of the CAP256-VRC26 lineage.

Rapid development of CAP256-VRC26.01

To gain insight into the development of V1V2-directed neutralization, we focused on the early antibody CAP256-VRC26.01, isolated at week 59, which neutralized 30% of clade C viruses and showed cross-clade neutralization of nearly 20% (Supplementary Fig. 1). Notably, this week 59 time point was 44 weeks after superinfection and only 21 weeks after the CAP256-VRC26 lineage was first detected by NGS. We also inferred heavy and light chains for two developmental intermediates (VRC26-I1 and VRC26-I2) (Fig. 6a and Extended Data Fig. 1) and characterized their function along with the UCA (Fig. 6b–e). The UCA bound and neutralized the superinfecting virus weakly, but did not bind or neutralize heterologous viruses. VRC26-I1, VRC26-I2 and CAP256-VRC26.01 demonstrated progressively greater binding and neutralization, with VRC26-I1 neutralizing 2 of 7 strains and VRC26-I2 neutralizing 6 of 7 strains (Fig. 6e), with dependence on residues in V2 (Fig. 6c). Interestingly, the primary infecting virus was neither bound nor neutralized by the UCA, intermediates, or CAP256-VRC26.01 (Fig. 6c and Supplementary Fig. 11). These data provide further evidence that the CAP256-VRC26 lineage was initiated by interaction with a superinfecting-virus-like V1V2. Subsequent affinity maturation, focused within CDR H3 (Fig. 6f and Extended Data Table 1), allowed for progressively greater binding and neutralization with increased viral diversity preceding the emergence of neutralization breadth. On the basis of the inferred UCA, CAP256-VRC26.01 diverged 11% from germline heavy chain and 7% from germline light chain (Fig. 6f). Thus, once an appropriate gene recombination

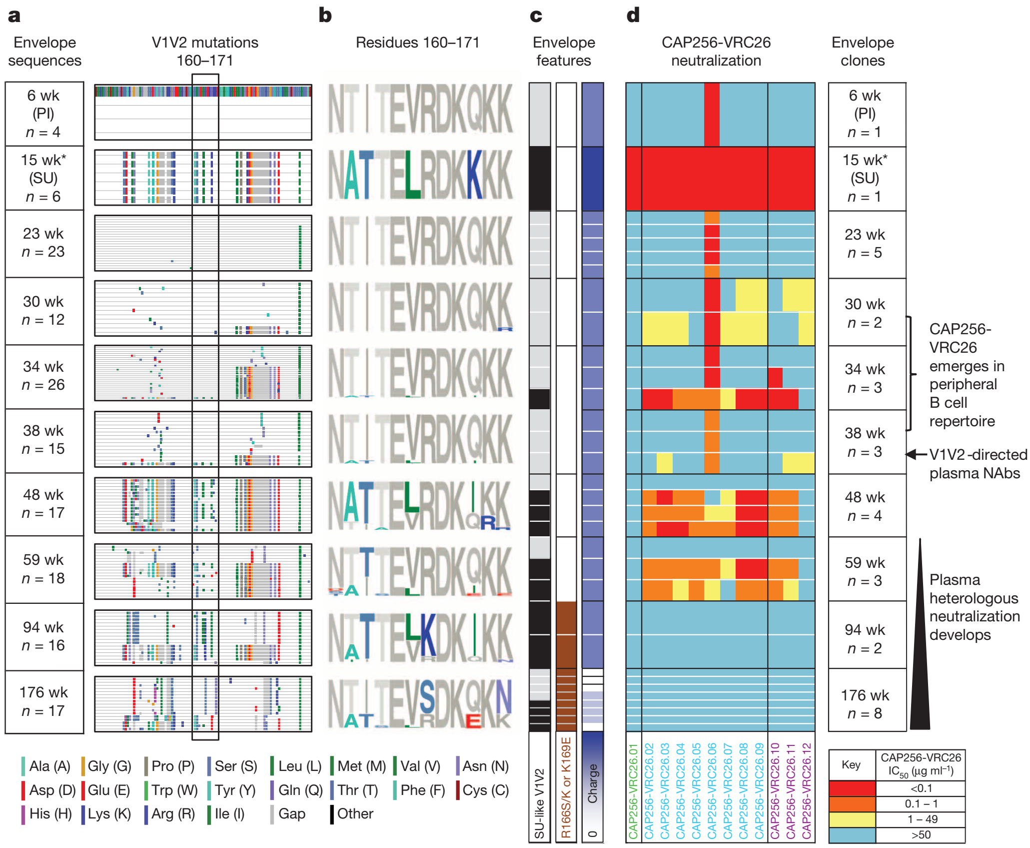


Figure 5 | HIV-1 Env evolution and the development of the CAP256-VRC26 lineage. a, V1V2 sequences are shown in highlighter format with the primary infecting virus (PI) designated as master and V2 residues 160 to 171 boxed. Asterisk at week 15 denotes sequences amplified with strain-specific primers matching the superinfecting virus (SU) virus. **b**, Logogram of the V2 epitope for all CAP256 sequences, with mutations away from the PI (master sequence) in colour. **c**, SU-like V1V2 sequences are indicated by black (present) and grey (absent) boxes. Escape mutations (K169E or R166S/K) are indicated by brown boxes. The net charge of the V2 epitope (residues 160 to 171) is shown in purple/white, ranging from +3 to 0. White lines separate clones within a time point; black lines separate time points. **d**, Neutralization by the 12 CAP256-VRC26 monoclonal antibodies of representative longitudinal Env clones isolated between 6 and 176 weeks post-infection (weeks shown at far right). The CAP256 monoclonal antibodies are coloured by time of isolation (as in Fig. 1). The development of the CAP256-VRC26 antibody lineage, V1V2-directed plasma neutralizing antibodies and plasma heterologous neutralization, are indicated on the right.

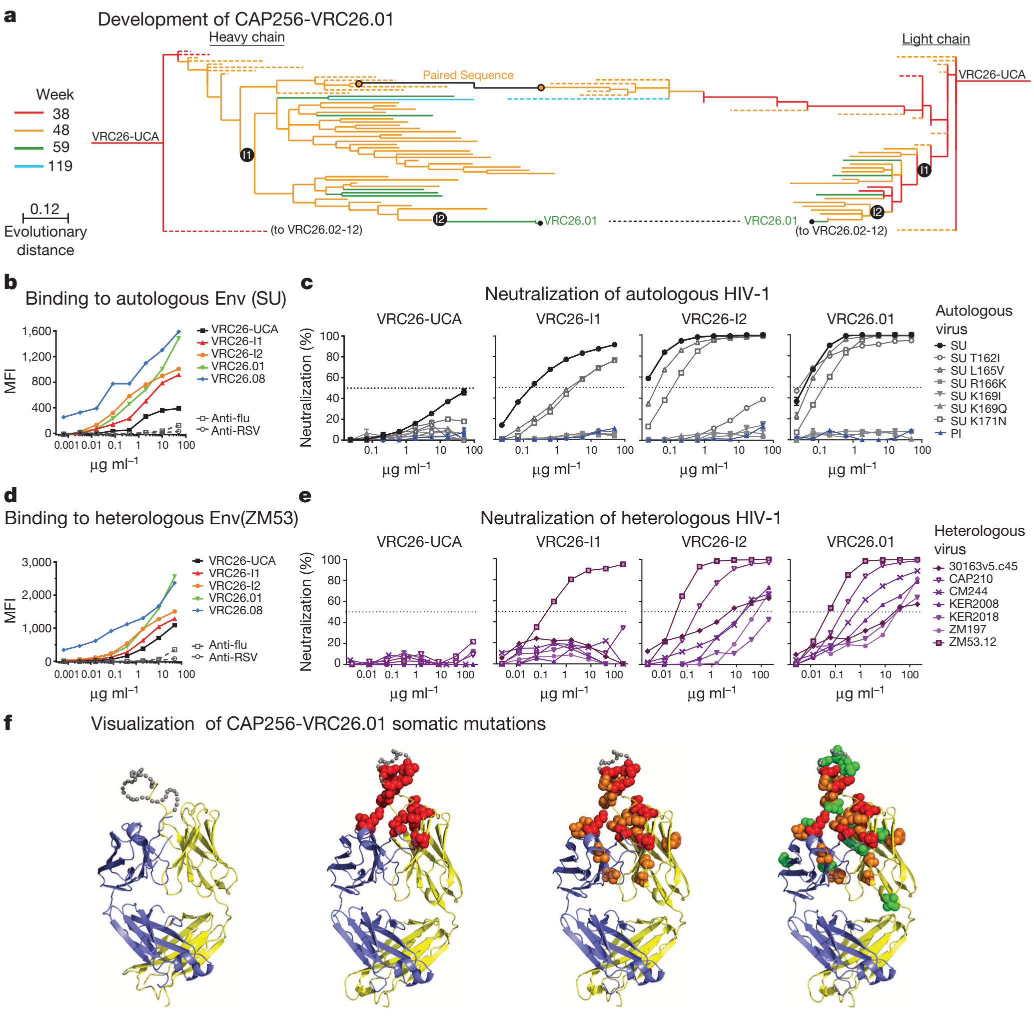
allows for B-cell receptor recognition of the trimeric V1V2 epitope, development of cross-reactive neutralization can be achieved with moderate somatic mutation in a matter of months.

Vaccine implications

The V1V2 region of HIV-1 is a common target of serum neutralizing antibodies[6,16–18]. In the RV144 Thai vaccine trial, an increased level of binding antibodies to the V1V2 region was associated with a reduced risk of infection[36] and viral sieve analysis showed immune pressure in the same region[37]. Although the vaccine in the RV144 trial did not elicit broadly neutralizing V1V2-directed antibodies similar to those described here and elsewhere[19–21], a more effective vaccine would ideally elicit cross-reactive neutralizing antibodies[1–3,38]. Previously described V1V2 neutralizing antibodies, and the CAP256-VRC26 lineage, all have long CDR H3 regions that are necessary to penetrate the glycan shield and engage a V1V2 epitope (Extended Data Table 1). An important unanswered question has been whether these long CDR H3s are fully formed by VDJ recombination, as has been seen in HIV-uninfected donors[39], or emerge by insertions during the process of affinity maturation. We show here that the 35-residue CDR H3 of the CAP256-VRC26 UCA was produced during initial gene rearrangement and therefore existed at the level of the naive B cell receptor.

A potential rate-limiting developmental step in the CAP256-VRC26 lineage is the gene rearrangement that generated its UCA. By one estimate, human B cells with recombined antibody genes encoding long (≥24 amino acids, international immunogenetics database (IMGT)[40] definition) or very long (≥28 amino acids) CDR H3s constitute ∼3.5% and 0.4%, respectively, of naive B cells[39]. These long B cell receptors have been associated with autoreactivity, and are subject to both central and peripheral deletion, resulting in an even smaller population of IgG$^+$ memory B cells[39,41]. We therefore tested the UCA and all 12 CAP256-VRC26 cloned antibodies for autoreactivity[42]. The UCA and mature CAP256-VRC26 antibodies demonstrated little or no reactivity with Hep2 cells or with cardiolipin (Extended Data Fig. 6b, c). In addition, NGS of CAP256 peripheral B cells indicated that <0.4% of sequences had CDR H3s of ≥ 28 amino acids (Extended Data Fig. 6d) suggesting that this donor did not have an unusually high frequency of clonal lineages with long CDR H3 regions.

We also inferred the virological events leading to the stimulation and evolution of the CAP256-VRC26 lineage by the superinfecting virus. Similar to the CH103 CD4-binding site lineage in donor CH505 (ref. 15), the autologous virus in CAP256 showed extensive diversification before the development of breadth. Subsequent antibody–virus interactions appeared to drive somatic mutation and development of cross-reactive neutralization. Finally, the ontogeny of V1V2-directed neutralizing antibodies revealed by the CAP256-VRC26 lineage indicates that neutralization potency and breadth can be achieved without extraordinary levels of somatic hypermutation. Although some neutralizing antibodies appear to require years of maturation[1,3,43,44], we show that a V1V2-directed B cell lineage can acquire HIV-1 neutralization breadth within months rather than years. The critical event appears to be an uncommon gene rearrangement that produces a B-cell receptor



	VRC26-UCA VH	VRC26-UCA Full UCA-HC	VRC26-UCA VL	VRC26-UCA Full UCA-LC	VRC26-I1 VH	VRC26-I1 Full UCA-HC	VRC26-I1 VL	VRC26-I1 Full UCA-LC	VRC26-I2 VH	VRC26-I2 Full UCA-HC	VRC26-I2 VL	VRC26-I2 Full UCA-LC	VRC26.01 VH	VRC26.01 Full UCA-HC	VRC26.01 VL	VRC26.01 Full-UCA-LC
Mutations (nt %)	0	0	0	0	2	3	2	2	6	7	3	4	8	11	4	7
(aa %)	0	0	0	0	6	8	3	3	11	14	5	5	16	21	7	8

Figure 6 | Development from UCA to CAP256-VRC26.01. **a**, Expanded view of the phylogenetic trees from Fig. 3c, highlighting the maturation pathway of CAP256-VRC26.01. Off-pathway branches were collapsed and are shown as dashed lines. Inferred intermediates VRC26-I1 and VRC26-I2 were expressed for functional analyses. **b**–**e**, Binding and neutralization of antibodies UCA, VRC26-I1, VRC26-I2, VRC26.01. **b**, **d**, Binding to cell-surface expressed Env (SU and ZM53). MFI, median fluorescence intensity. **c**, **e**, Neutralization of PI, SU and point mutants (**c**) and seven heterologous viruses (**e**). Bars, standard error of the mean (triplicates). **f**, Structural models of VRC26.01 lineage antibodies. Affinity matured residues are shown as spheres coloured according to the intermediate at which they first appear: red, VRC26-I1; orange, VRC26-I2; green, VRC26.01. Grey dots, disordered residues in the CDR H3. The number of changes from the UCA to each intermediate are noted for *V* gene only (VH or VL), or from the full UCA (UCA-HC or UCA-LC).

with a protruding, tyrosine-sulphated, anionic CDR H3. Identifying features of antigens able to engage naive B cells with such CDR H3s is a critical step in design of vaccines targeting V1V2. Such antigens could be screened for binding to the UCA versions of neutralizing antibodies as an indicator of the ability to engage an appropriate naive B cell receptor. This work also suggests that although an appropriate trimeric V1V2 construct may elicit neutralizing V1V2 antibodies, sequential immunogens that mirror viral evolution may be needed to drive the development of breadth. Overall, the precise delineation of the developmental pathway for the CAP256-VRC26 lineage should provide a basis for attempts to elicit broad V1V2-directed HIV-1-neutralizing antibodies.

METHODS SUMMARY

Serial blood samples were collected from HIV-1-infected subject CAP256 from 6 to 225 weeks after infection. Monoclonal antibodies CAP256-VRC26.01-12 were generated by single-B cell culture, microneutralization screening, RT–PCR, subcloning, and expression as described in[27,28,45]. CDR lengths used Kabat notation[29] except as indicated. Binding of CAP256-VRC26 antibodies to virus-like particles was assessed by ELISA[31] and binding to cell-surface expressed Env was measured by flow cytometry[30]. HIV-1 neutralizing activity of patient plasma and monoclonal antibodies was determined with Env-pseudoviruses using the TZM-bl cell line[46,47]. Neutralization fingerprints are the rank-order of neutralization potencies for an antibody against a set of diverse viral strains, calculated as in ref. 18. A 28Å reconstruction of the BG505 SOSIP.664 gp140 trimer with a single VRC26.09 Fab was obtained by negative stain EM using Appion, Xmipp, IMAGIC, and EMAN software. 454 pyrosequencing was performed as previously described[12,14] on samples from 8 time points after HIV-1 infection. High-throughput VH:VL pairing of peripheral blood CD27^{+} B cells was performed in single cell emulsions generated using a flow focusing apparatus[35]. Phylogenetic analysis, inference of UCA, and identification, synthesis, and expression of clone members were performed as described in the Methods. Epitope mapping onto the spike trimer was performed with the software package UCSF Chimera, using experimental data as described in Methods. Crystallographic analysis of Fab fragments was performed as described in the Methods. Structure modelling of disordered residues in Fab crystal structures was performed using Loopy software. Single-genome amplification and expression of *env* genes was performed as described in Methods and in refs 48, 49.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 September 2013; accepted 16 January 2014.
Published online 2 March 2014.

1. Haynes, B. F., Kelsoe, G., Harrison, S. C. & Kepler, T. B. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnol.* **30,** 423–433 (2012).
2. Kong, L. & Sattentau, Q. J. Antigenicity and immunogenicity in HIV-1 antibody-based vaccine design. *J. AIDS Clinic. Res.* (Suppl. 8) 003 (2012).
3. Mascola, J. R. & Haynes, B. F. HIV-1 neutralizing antibodies: understanding nature's pathways. *Immunol. Rev.* **254,** 225–244 (2013).
4. Richman, D. D., Wrin, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **100,** 4144–4149 (2003).
5. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422,** 307–312 (2003).

6. Gray, E. S. *et al.* The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4$^+$ T cell decline and high viral load during acute infection. *J. Virol.* **85,** 4828–4840 (2011).
7. Piantadosi, A. *et al.* Breadth of neutralizing antibody response to human immunodeficiency virus type 1 is affected by factors early in infection but does not influence disease progression. *J. Virol.* **83,** 10269–10274 (2009).
8. Sather, D. N. *et al.* Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection. *J. Virol.* **83,** 757–769 (2009).
9. Doria-Rose, N. A. *et al.* Breadth of human immunodeficiency virus-specific neutralizing activity in sera: clustering analysis and association with clinical variables. *J. Virol.* **84,** 1631–1636 (2010).
10. Glanville, J. *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc. Natl Acad. Sci. USA* **106,** 20216–20221 (2009).
11. Briney, B. S., Willis, J. R., McKinney, B. A. & Crowe, J. E. Jr. High-throughput antibody sequencing reveals genetic evidence of global regulation of the naive and memory repertoires that extends across individuals. *Genes Immun.* **13,** 469–473 (2012).
12. Wu, X. *et al.* Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333,** 1593–1602 (2011).
13. Zhu, J. *et al.* Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc. Natl Acad. Sci. USA* **110,** 6470–6475 (2013).
14. Zhu, J. *et al.* Somatic populations of PGT135–137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. *Front. Microbiol.* **3,** 315 (2012).
15. Liao, H. X. *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496,** 469–476 (2013).
16. Walker, L. M. *et al.* A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. *PLoS Pathog.* **6,** e1001028 (2010).
17. Lynch, R. M. *et al.* The B cell response is redundant and highly focused on V1V2 during early subtype C infection in a Zambian seroconverter. *J. Virol.* **85,** 905–915 (2011).
18. Georgiev, I. S. *et al.* Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. *Science* **340,** 751–756 (2013).
19. Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326,** 285–289 (2009).
20. Bonsignori, M. *et al.* Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J. Virol.* **85,** 9998–10009 (2011).
21. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477,** 466–470 (2011).
22. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480,** 336–343 (2011).
23. Pancera, M. *et al.* Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1–V2-directed antibody PG16. *Nature Struct. Mol. Biol.* **20,** 804–813 (2013).
24. Julien, J. P. *et al.* Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. *Proc. Natl Acad. Sci. USA* **110,** 4351–4356 (2013).
25. Moore, P. L. *et al.* Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. *J. Virol.* **85,** 3128–3141 (2011).
26. Moore, P. L. *et al.* Multiple pathways of escape from HIV broadly cross-neutralizing V2-dependent antibodies. *J. Virol.* **87,** 4882–4894 (2013).
27. Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491,** 406–412 (2012).
28. Tiller, T. *et al.* Efficient generation of monoclonal antibodies from single human B cells by single cell RT–PCR and expression vector cloning. *J. Immunol. Methods* **329,** 112–124 (2008).
29. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. *Sequences of Proteins of Immunological Interest* (U.S. Department of Health and Human Services, National Institutes of Health, 1991).
30. Pancera, M. & Wyatt, R. Selective recognition of oligomeric HIV-1 primary isolate envelope glycoproteins by potently neutralizing ligands requires efficient precursor cleavage. *Virology* **332,** 145–156 (2005).
31. Tong, T., Crooks, E. T., Osawa, K. & Binley, J. M. HIV-1 virus-like particles bearing pure env trimers expose neutralizing epitopes but occlude nonneutralizing epitopes. *J. Virol.* **86,** 3574–3587 (2012).
32. Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342,** 1484–1490 (2013).
33. Julien, J. P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342,** 1477–1483 (2013).
34. Doria-Rose, N. A. *et al.* A short segment of the HIV-1 gp120 V1/V2 region is a major determinant of resistance to V1/V2 neutralizing antibodies. *J. Virol.* **86,** 8319–8323 (2012).
35. DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature Biotechnol.* **31,** 166–169 (2013).
36. Haynes, B. F. *et al.* Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N. Engl. J. Med.* **366,** 1275–1286 (2012).
37. Rolland, M. *et al.* Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature* **490,** 417–420 (2012).
38. Overbaugh, J. & Morris, L. The antibody response against HIV-1. *Cold Spring Harb. Perspect. Med.* **2,** a007039 (2012).
39. Briney, B. S., Willis, J. R. & Crowe, J. E. Jr. Human peripheral blood antibodies with long HCDR3s are established primarily at original recombination using a limited subset of germline genes. *PLoS ONE* **7,** e36750 (2012).
40. Lefranc, M. P. *et al.* IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* **27,** 55–77 (2003).
41. Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301,** 1374–1377 (2003).
42. Haynes, B. F. *et al.* Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* **308,** 1906–1908 (2005).
43. Kwong, P. D. & Mascola, J. R. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* **37,** 412–425 (2012).
44. Burton, D. R. *et al.* A blueprint for HIV vaccine discovery. *Cell Host Microbe* **12,** 396–407 (2012).
45. Huang, J. *et al.* Isolation of human monoclonal antibodies from peripheral blood B cells. *Nature Protocols* **8,** 1907–1915 (2013).
46. Shu, Y. *et al.* Efficient protein boosting after plasmid DNA or recombinant adenovirus immunization with HIV-1 vaccine constructs. *Vaccine* **25,** 1398–1408 (2007).
47. Montefiori, D. C. Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol. Biol.* **485,** 395–405 (2009).
48. Kraus, M. H. *et al.* A *rev1-vpu* polymorphism unique to HIV-1 subtype A and C strains impairs envelope glycoprotein expression from *rev-vpu-env* cassettes and reduces virion infectivity in pseudotyping assays. *Virology* **397,** 346–357 (2010).
49. Salazar-Gonzalez, J. F. *et al.* Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J. Virol.* **82,** 3952–3970 (2008).
50. Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G. & Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455,** 109–113 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the participants in the CAPRISA 002 study for their commitment. For technical assistance and advice, we thank: K. Mlisana, S. Sibeko, N. Naicker, the CAPRISA 002 clinical team, N. Samsunder, S. Heeralall, B. Lambson, M. Madzivhandila, T. Khoza, C. Mitchell Scheepers, E. Turk, C.-L. Lin, M. Roederer, J. Stuckey, B. Hartman, G. Loots, J. H. Lee, G. Ippolito, B. Briney, S. Hunicke-Smith and J. Wheeler, and members of the WCMC HIVRAD Core and the NIH Vaccine Research Center HIMS, HIMC, SBS and SBIS sections. We thank J. Baalwa, D. Ellenberger, F. Gao, B. Hahn, K. Hong, J. Kim, F. McCutchan, D. Montefiori, J. Overbaugh, E. Sanders-Buell, G. Shaw, R. Swanstrom, M. Thomson, S. Tovanabutra and L. Zhang for contributing the HIV-1 Envelope plasmids used in our neutralization panel. Funding was provided by the intramural research programs of the Vaccine Research Center and NIAID, the Fogarty International Center, NHGRI, and NIGMS of the National Institutes of Health, USA; the International AIDS Vaccine Initiative; the National Science Foundation; Scripps CHAV-ID; the South African Department of Science and Technology; and fellowships from the Wellcome Trust, Hertz Foundation, Donald D. Harrington Foundation, Poliomyelitis Research Foundation and the National Research Foundation of South Africa. Use of sector 22 (Southeast Region Collaborative Access team) at the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract number W-31-109-Eng-38.

Author Contributions N.A.D.-R., C.A.S., J.G. and P.L.M. contributed equally to this work. N.A.D.-R., C.A.S., J.G., P.L.M. and J.N.B., designed and performed experiments, analysed data and wrote the manuscript. L.M., P.D.K., L.S. and J.R.M. conceived and designed the experiments, analysed data, and wrote the manuscript. B.J.D., M.J.E., I.S.G, H.J.K., M.P. and R.P.S. conducted experiments and analysed data. H.R.A.-T., B.T.B., E.T.C., A.C., K.H.H., R.K., M.K.L., K.M., M.N., S.O., Ry.S.R., Re.S.R., S.D.S., C.K.W., Y.Y., J.C.M. and NISC conducted experiments. C.W. and A.D. contributed analysis tools and data analysis. S.S.A.K. and N.J.G conceived and managed the CAPRISA cohorts. J.M.B., R.W.S., I.A.W., J.P.M., A.B.W., G.G., N.S.L., D.J.S., C.S. and Z.Z. analysed data.

Author Information Coordinates and structure factors for CAP256-VRC26 lineage Fabs have been deposited with the Protein Data Bank under accession codes 4ODH, 4OCR, 4OD1, 4ORG, 4OCW, 4OD3 and 4OCS. The EM reconstruction density for the CAP256-VRC26.09 complex with BG505 SOSIP.664 trimer has been deposited with the Electron Microscopy Data Bank under accession code EMD-5856. We have also deposited deep sequencing data used in this study to National Center for Biotechnology Information Short Reads Archives (SRA) under accession numbers SRP034555 and SRP017087. Information deposited with GenBank includes: the heavy- and light-chain variable region sequences of cloned antibodies CAP256-VRC26.01-12, UCA, I1 and I2 (accession numbers KJ134860–KJ134889); bioinformatically identified VRC26-related sequences from B cell transcripts: 680 heavy chains and 472 light chains (accession numbers KJ133708 – KJ134387, KJ134388 – KJ134859); and CAP256 Env sequences (accession numbers KF996576 – KF996716). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed for CAPRISA and viral evolution to L.M. (lynnm@nicd.ac.za), for crystallography to P.D.K. (pdkwong@nih.gov), for NGS to L.S. (lss8@columbia.edu), and for isolated antibodies to J.R.M. (jrmascola@nih.gov).

METHODS

Study subject. CAPRISA participant CAP256 was enrolled into the CAPRISA acute infection study[51] that was established in 2004 in KwaZulu-Natal, South Africa for follow-up and subsequent identification of HIV seroconversion. CAP256 was one of the 7 women in this cohort who developed neutralization breadth[6]. The CAPRISA 002 acute infection study was reviewed and approved by the research ethics committees of the University of KwaZulu-Natal (E013/04), the University of Cape Town (025/2004), and the University of the Witwatersrand (MM040202). CAP256 provided written informed consent for study participation. Samples were drawn between 2005–09.

Isolation and expression of CAP256-VRC26 family genes. PBMC isolated from CAP256 blood draws at weeks 59, 119 and 206 were stained and sorted for IgG^{+} B cells on a FACS Aria II as described in ref. 18. Cells were plated at two B cells per well in 384-well plates and cultured for 14 days in the presence of IL-2, IL-21, and CD40L-expressing irradiated feeder cells, as described in refs 27 and 45. Culture supernatants were screened by microneutralization as described in ref. 52 against HIV-1 ZM53.12 and either CAP45.G3 or CAP210.E8 Env-pseudoviruses. Kappa and lambda light chain gene and IgG heavy chain gene variable regions were amplified from neutralization-positive wells, subcloned, expressed and purified as described in ref. 18. Heavy chains were reconstituted as IgG1. The efficiency of cloning was as follows. For week 59, a total of 15,000 B cells (7,500 wells) were plated, 8.3% of wells produced IgG, 4 were positive in microneutralization, and one heavy-light chain pair was recovered. For week 119, a total of 45,000 B cells were plated, 48% of wells produced IgG, 49 wells were positive in microneutralization, and 8 heavy-light chain pairs were recovered. For week 206, a total of 42,000 B cells were plated, 29% of wells produced IgG, 34 wells were positive in microneutralization and 3 heavy-light chain pairs were recovered.

The antibodies are numbered CAP256-VRC26.01-.12 in order of the time point of the sample from which they were isolated, and then the degree of heavy-chain somatic mutation.

Neutralization assays. Single round of replication Env-pseudoviruses were prepared, titred and used to infect TZM-bl target cells as described previously[46,47]. Neutralization breadth of CAP256-VRC26.01, .03, .06, and .08 were determined using a previously described[18,53] panel of 194 geographically and genetically diverse Env-pseudoviruses representing the major subtypes and circulating recombinant forms. The remaining antibodies were assayed on a subset of this panel. The data were calculated as a reduction in luminescence units compared with control wells, and reported as half-maximum inhibitory concentration (IC_{50}) in micrograms per microlitre for monoclonal antibodies, or reciprocal dilution (ID_{50}) for plasma samples.

Neutralization fingerprints. Owing to the high sequence variability of HIV-1 Env, different viral strains may exhibit different neutralization sensitivities to the same antibody, and this pattern of neutralization variation can be used to define the neutralization fingerprint for a given antibody. Namely, the neutralization fingerprint of an antibody is defined as the rank-order of neutralization potencies for the antibody against a set of diverse viral strains[18].

The correlations between the neutralization fingerprints of the CAP256-VRC26 antibodies and the neutralization patterns of four longitudinal serum time points (at 59, 106, 159, and 220 weeks post-infection) were computed over a set of 29 HIV-1 strains (6535.3, AC10.29, CAAN.A2, CAP210.E8, CAP244.D3, CAP45.G3, DU156.12, DU172.17, DU422.01, PVO.04, Q168.a2, Q23.17, Q259.d2.17, Q461.e2, Q769.d22, Q842.d12, QH0692.42, REJO.67, RHPA.7, SC422.8, THRO.18, TRJO.58, TRO.11, WITO.33, ZM109.4, ZM135.10a, ZM197.7, ZM233.6, ZM53.12)[18]. The correlations between the neutralization potencies of the CAP256-VRC26 antibodies and a reference set of antibodies targeting the four major sites of vulnerability, with at most two antibodies per unique donor, were computed over a set of 41 HIV-1 strains (6535.3, 0260.v5.c36, 6405.v4.c34, AC10.29, C1080.c3, CAAN.A2, CAP210.E8, CAP244.D3, CAP45.G3, CNE3, DU156.12, DU172.17, DU422.01, KER2008.12, KER2018.11, MB201.A1, MB539.2B7, PVO.04, Q168.a2, Q23.17, Q259.17, Q461.e2, Q769.d22, Q842.d12, QH0692.42, REJO.67, RHPA.7, RW020.2, SC422.8, TH976.17, THRO.18, TRJO.58, TRO.11, UG037.8, WITO.33, ZM109.4, ZM135.10a, ZM197.7, ZM214.15, ZM249.1, ZM53.12). The correlations between the neutralization patterns of the four longitudinal serum time points and the neutralization fingerprints of the reference antibodies were computed over a set of 28 HIV-1 strains (6535.3, AC10.29, CAAN.A2, CAP210.E8, CAP244.D3, CAP45.G3, DU156.12, DU172.17, DU422.01, PVO.04, Q168.a2, Q23.17, Q259.17, Q461.e2, Q769.d22, Q842.d12, QH0692.42, REJO.67, RHPA.7, SC422.8, THRO.18, TRJO.58, TRO.11, WITO.33, ZM109.4, ZM135.10a, ZM197.7, ZM53.12). For the reference antibodies, data from multiple neutralization experiments were averaged and consolidated. All correlations are based on the Spearman's rank correlation coefficient.

Virus-like particle ELISA. VLP ELISAs were performed as described previously[31]. Briefly, VLPs were produced by PEI-based cotransfection of 293T cells with a pCAGGS-based, Env-expressing plasmid and the Env-deficient HIV-1 genomic backbone plasmid pNL-LucR-E. VLPs were coated on ELISA wells at 20× the concentration in transfection supernatants. Monoclonal antibody binding was then assessed by ELISA, omitting detergent in PBS wash buffers and probing with an anti-human Fc alkaline phosphatase conjugate (Accurate, Westbury, NY) and SigmaFAST p-nitrophenyl phosphate tablets (Sigma). Plates were read at 405 nm.

Cell-surface Env binding. 293T cells were transiently transfected with plasmids encoding Env ZM53.12 or CAP256-SU with deletions of the cytoplasmic tail[30]. For binding experiments: after 2 days, the cells were stained with ViVid viability dye (Invitrogen) followed by serial dilutions of antibodies, two washes with PBS/5% FBS, then R-PE-conjugated F(ab) goat anti-human IgG specific for the Fc fragment (Jackson ImmunoResearch) at a 1:200 dilution[54]. For competition assays, the cells were stained with ViVid viability dye followed by biotinylated CAP256-VRC26.08 ($0.8\,\mu g\,ml^{-1}$) premixed with serially diluted unlabelled competitor antibodies. After incubation and 2 washes, cells were stained with streptavidin-PE (Invitrogen) at 1:200 dilution. Cells were analysed on a BD LSRII (Becton Dickinson). Binding was measured as the median fluorescence intensity (MFI) for each sample minus the MFI of cells stained with secondary antibody only.

Polyreactivity analysis of antibodies. Antibody binding to cardiolipin was determined as in ref. 42. Briefly, using the QUANTA Lite ACA IgG III ELISA kit (Zeus Scientific) per manufacturer's protocol, each antibody was diluted to $100\,\mu g\,ml^{-1}$ in the kit sample diluent and tested in threefold serial dilutions. Results shown are representative of at least two independent ELISAs. Positive and negative controls were included on each plate, and values three times above background were considered positive. Antibody reactivity to a human epithelial cell line (HEp-2) was determined with the ANA/HEp-2 Cell Culture IFA Test System (Zeus Scientific) per manufacturer's protocol, as described in ref. 42. Antibodies were diluted to $50\,\mu g\,ml^{-1}$ and $25\,\mu g\,ml^{-1}$ in ZOBRA-NS diluent. Positive and negative controls were included on each slide. Antibodies were scored negative, indeterminate, or positive (1+ to 4+) at each dilution. Results are representative of at least two independent experiments.

Electron microscopy (EM) and image processing. VRC26.09 Fabs in complex with BG505 SOSIP.664 gp140 trimer produced in HEK 293S cells were analysed by negative stain EM. A 3 μl aliquot of $\sim 8\,\mu g\,ml^{-1}$ of the complex was applied for 15 s onto a glow discharged, carbon-coated 400 Cu mesh grid and stained with 2% uranyl formate for 20 s. Grids were imaged using a FEI Tecnai T12 electron microscope operating at 120 kV using a 52,000× magnification and electron dose of $25\,e^{-}/\text{Å}^2$, which resulted in a pixel size of 2.05Å at the specimen plane. Images were acquired with a Tietz 4k × 4k CCD camera in 5° tilt increments from 0° to 50° at a defocus of 1,000 nm using LEGINON[55].

Particles were picked automatically by using DoG Picker and put into a particle stack using the Appion software package[56,57]. Initial reference free 2D class averages were calculated using particles binned by 2 via the Xmipp Clustering 2D Alignment and sorted into 128 classes[58]. Particles corresponding to the complexes were selected into a substack and another round of reference free alignment was carried out with unbinned particles using Xmipp Clustering 2D alignment and IMAGIC softwares[59]. To generate an *ab initio* 3D starting model, a template stack of 44 images of 2D class averages was used without imposing symmetry. The resulting starting model was refined against 2D class averages for 9 cycles and subsequently with 6,763 raw particles for 9 cycles using EMAN[60]. The resolution of the final reconstruction was calculated to be 28Å using an FSC cut-off value of 0.5.

High-throughput sequencing. Amplicon for 454 next-generation sequencing was prepared as described[12,14] with slight modifications as indicated. Briefly, mRNA was prepared from 10–15 million PBMC using an Oligotex kit (Qiagen). cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen) and oligo-dT(12–18) primers. Individual PCR reactions were performed with Phusion polymerase for 30 cycles. Primers (Supplementary Table 4) consisted of pools of 5–7 oligonucleotides specific for all lambda gene families or VH3 family genes, and had adapters for 454 next generation sequencing. For week 176 only, heavy-chain PCR was performed with primers for all VH families, and mixed lambda and kappa primers were used for light chain (Supplementary Table 4). PCR products were gel-purified (Qiagen). Pyrosequencing of the PCR products was performed on a GSFLX sequencing instrument (Roche-454 Life Sciences, Bradford, CT, USA) on a half chip per reaction (full chips for week 176). On average, ~250,000 raw reads were produced.

High-throughput linkage of VH and VL transcripts was performed in single cell emulsions generated using a flow focusing apparatus[35] (B.J.D., manuscript in preparation). $CD27^{+}$ B cells were isolated from CAP256 PBMCs collected at 34, 48, 59, 69, and 119 weeks post-infection by magnetic bead sorting (Miltenyi Biotec, Auburn, CA). Cells from weeks 34 and 119 were divided in two groups and half of the cells were analysed with FR1 primers[35], while the other half were analysed with leader peptide primers[41] (Supplementary Table 5). All other time points were analysed in a single group using only FR1 primers (Supplementary Table 1). Overlap extension RT–PCR was performed as previously reported[35], with extension time increased to 125 s. Nested PCR was performed as described previously with a 23-s

extension time and PCR products were sequenced using the Illumina 2×250 bp MiSeq platform. Raw reads were quality-filtered for an Illumina Q-score of 20 in 50% of bases. VRC26-class VH and paired VL sequences were identified via BLAST against CDR-H3 nucleotide sequences of the 12 culture-isolated antibodies.

Antibodyomics pipeline. Raw 454 data was processed using a pipeline implemented in Python, similar to one we reported previously[14]. Briefly, reads were filtered for length, keeping only those between 300 and 600 nucleotides. Germline V genes were then assigned to each read using BLAST with empirically optimized parameters. Reads for which no V gene match was found with an *e*-value $\leq 10^{-10}$ were discarded. For reads assigned to any VH3-30 or Vλ1-51 allele, (the CAP256-VRC26 germline genes), ClustalW2 (ref. 61) was used to calculate the sequence identity to the germline and each isolated antibody. These data were plotted as density heat maps using ggplot2 in R to produce identity-divergence plots (Fig. 3b and Supplementary Fig. 4).

Finding clonally related sequences. Reads that were assigned to the same V genes as CAP256-VRC26, VH3-30 and Vλ1-51, were submitted to IMGT High-Vquest[62] (http://www.imgt.org/IMGTindex/IMGTHighV-QUEST.html), and the results, including automated sequence corrections, were used to further sieve for lineage-related sequences. Reads assigned to *J* genes matching CAP256-VRC26 (JH3 or Jλ1), and having similar divergence (± 15%) in the *V* and *J* genes, similar (± 10%) nucleotide and amino acid divergences in the *V* gene, and containing a continuous open reading frame throughout the entire variable region, were selected for further processing. Next, reads from all time points were pooled and clustered at 97.25% sequence identity (twice the standard deviation of expected 454 sequencing error)[14] using CDHit[63]. For each cluster, a representative sequence was chosen from the earliest possible time point. The choice of cluster representatives from the earliest time points at which they appeared was critical to maintaining information on the chronology of lineage development in subsequent analyses. This procedure yielded 8,485 unique heavy chain and 6,410 unique light chain sequences.

To identify CAP256-VRC26 lineage-member heavy chains, we performed intra-donor phylogenetic analysis[14] on the unique 454 sequence set using the heavy chain sequences of the 12 isolated CAP256-VRC26 antibodies. 707 sequences were identified as likely lineage members, of which 27 were discarded after manual inspection, resulting in a total of 680 unique CAP256-VRC26 lineage heavy chain sequences.

To identify light chain lineage members, a sieve requiring at least 92% sequence identity in CDR L3 to one of the isolated antibodies resulted in 495 sequences. Joinsolver[64] was used to examine the *V-J* junctions of these sequences in detail, to ensure that the recombination points matched those known for the isolated antibodies (Supplementary Fig. 5). This gave a total of 472 unique CAP256-VRC26 lineage light chain sequences.

Paired reads that were identified as members of the CAP256-VRC26 lineage were clustered using CDHit[63] at 95% sequence identity and consensus VH and VL sequences were generated for each cluster containing two or more pairs. Blast was then used to align the resulting sequences to all clonally related sequences identified from the 454 sequencing as described above. Gapless alignments covering at least 190 nucleotides at 97% or greater sequence identity were considered to be matches. Two of the 157 paired sequences determined to be members of the CAP256-VRC26 lineage matched known CAP256-VRC26 lineage sequences in both VH and VL 454 data sets. An additional 4 VH sequences and 1 VL sequence were found in the 454 data, but their light or heavy chain partners were not present.

Computation of phylogenetic trees. Phylogenetic trees were constructed from 454 data and the sequences of antibodies isolated from B cell culture. Raw data are shown in Nexus format in Supplementary Figs 12 and 13. MEGA5 (ref. 65) was used to select the general time-reversible model with a gamma-distributed rate parameter (GTR+G)[66] as the best mathematical model for building a maximum-likelihood tree from the CAP256-VRC26 lineage sequences. FASTML[67] was then used to estimate the gamma parameter and build separate maximum likelihood trees for heavy and light chain sequences (including the isolated antibodies) and these were rooted on the germline *V* gene sequences. Two branches of the light chain tree were manually moved to match their positioning in the heavy chain tree based on the evidence from trees constructed solely with the 12 isolated antibodies. Analysis with DNAML from PHYLIP (Phylogeny Inference Package) version 3.6 (Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle) (http://cmgm.stanford.edu/phylip/dnaml.html) showed that these rearrangements did not significantly alter the log-likelihood score of the tree.

To create a condensed version of the heavy chain phylogenetic tree (Fig. 4c), CDR H3 sequences were clustered using a 95% sequence identity threshold and requiring that all CDR H3s in a cluster have the same length. Isolated antibodies and monophyletic clusters with at least five members were represented by a single leaf, while all other sequences were removed from the tree. In cases where an internal node was deleted, branch lengths above and below that node were summed, so that the tree depths of all remaining sequences were maintained.

UCA and inferred intermediates. The phylogenetic trees of all heavy and all light chain lineage members calculated above (Fig. 3c and Extended Data Fig. 1) were input into the DNAML maximum likelihood software package to infer ancestral sequences. These are a direct consequence of the input sequences and the mathematical model used to build the trees; the gamma distribution found by FASTML above was used and the topology of the tree was held fixed, so no further information was added. The calculated heavy chain UCA was identical to the germline VH3-30*18 allele. Although the VH3-30*03 allele is only one nucleotide different from *18, germline sequencing of this donor showed that she carries the *18 allele and not the *03 allele (Cathrine Mitchell Scheepers, personal communication). The inferred UCA is very similar to low-divergence sequences found in the week 38 data set (Extended Data Fig. 6).

To test intermediates in the development of CAP256-VRC26.01, two internal nodes were chosen from the phylogenetic trees to be approximately equally spaced in terms of evolutionary distance and the inferred sequences were retrieved using DNAML. Successful complementation of inferred heavy and light chains for each intermediate suggests that the lineage is well sampled by the 454 data and that the calculated phylogenetic trees successfully capture the coupled evolutionary dynamics of heavy and light chains.

Logograms for CDR H3s were made with Weblogo[68].

X-ray crystallography. VRC26.UCA Fab was prepared by digesting purified IgG with Lys-C at 37 °C for 2 h. The reaction was then quenched by the addition of cOmplete protease inhibitors (Roche). For VRC26.01, VRC26.03, VRC26.04, VRC26.06, VRC26.07 and VRC26.10 Fab preparation, an HRV3C recognition site (GLEVLFQGP) was inserted after Lys 235 and purified IgG was incubated with HRV3C protease overnight at 4 °C. For all, the digested antibodies were passed over Protein A agarose to remove the Fc fragment. The Fab was further purified over a Superdex 200 gel filtration column and concentrated aliquots were stored at −80 °C. All Fabs were screened against 576 crystallization conditions using a Cartesian Honeybee crystallization robot. Initial crystals were grown by the vapour diffusion method in sitting drops at 20 °C by mixing 0.2 μl of protein complex with 0.2 μl of reservoir solution. Crystals were manually reproduced in hanging drops by mixing 1.0 μl protein complex with 1.0 μl reservoir solution. VRC26-UCA was crystallized with a reservoir solution of 27% PEG 8000 and 0.1 M HEPES pH 7.5 and was flash frozen in liquid nitrogen with 20% PEG 400 as a cryoprotectant. VRC26.01 was crystallized with a reservoir solution of 32% PEG 400, 4% PEG 3350 and 0.1 M sodium acetate pH 5.5 and was flash frozen in liquid nitrogen with 20% ethylene glycol as a cryoprotectant. VRC26.03 was crystallized with a reservoir solution of 22% PEG 8000, 5% MPD and 0.1 M imidazole pH 6.5 and was flash frozen in liquid nitrogen with 20% xylitol as a cryoprotectant. VRC26.04 was crystallized with a reservoir solution of 14% PEG 3350, 25% ispropanol and 0.1 M Tris pH 8.5 and was flash frozen in liquid nitrogen with 20% ethylene glycol as a cryoprotectant. VRC26.06 was crystallized with a reservoir solution of 3 M sodium formate and 0.1 M Tris pH 7.5 and was flash frozen in liquid nitrogen with 20% xylitol as a cryoprotectant. VRC26.07 was crystallized with a reservoir solution of 4% PEG 8000, 0.1 M zinc acetate and 0.1 M MES pH 6 and was flash frozen in liquid nitrogen with 20% glycerol as a cryoprotectant. VRC26.10 was crystallized with a reservoir solution of 22% PEG 4000, 0.4M sodium acetate and 0.1 M Tris pH 7.5 and was flash frozen in liquid nitrogen with no cryoprotectant.

Data for all crystals were collected at a wavelength of 1.00Å at SER-CAT beamlines ID-22 and BM-22 (Advanced Photon Source, Argonne National Laboratory). All diffraction data were processed with the HKL2000 suite[69] and model building and refinement were performed in COOT[70] and PHENIX[71], respectively. For VRC26.03 Fab data, a molecular replacement solution consisting of one Fab molecule per asymmetric unit was obtained using PHASER with a search model from PDB ID 3F12. VRC26.03 then served as a search model for all remaining VRC26 Fabs. Throughout the refinement processes, a cross validation (Rfree) test set consisting of 5% of the data was used and hydrogen atoms were included in the refinement model. Structure validations were performed periodically during the model building/refinement process with MolProbity[72]. Ribbon diagram representations of protein crystal structures were made with PyMOL[73] and electrostatics were calculated and rendered with UCSF Chimera[74].

Structure modelling on trimers. Defined locations of the V1V2, V3-glycan and CD4-binding sites were mapped directly onto EM density of the unliganded HIV-1 BAL spike (EMD-5019)[50] using the software package UCSF Chimera[74]. The CD4-binding site was defined by aligning density of the VRC01-bound BAL spike (EMD-5457)[75] with the unliganded map and fitting a crystal structure of VRC01-bound gp120 (PDB accession number 3NGB)[76] to the density. EM density in close proximity to the Fab structure was colored to highlight the region of contact. The same procedure was used to define the V3-glycan region using a PGT128-bound trimer (EMD-1970) and crystal structure (PDB id 3TYG)[77] and the V1V2

region using the PG9-bound BG505 SOSIP trimer (EMD-2241)[24] and a crystal structure of V1V2-bound PG9 (PDB accession number 3U4E)[22]. The fit of the PG9-V1V2 crystal structure to the SOSIP trimer was used to model the trimeric orientation of V1V2 using the threefold symmetry of the HIV-1 spike. The BG505.664 SOSIP crystal structure[33], PDB 4NCO, was presented to highlight the quaternary location of V1V2 point mutations. Side chains of residues 166 and 167, not seen in the crystal structure, were modelled. The Man5 glycan at N160, also not seen in the crystal structure, is represented as in the crystal structure of the PG9-V1V2 complex (PDB accession code 3U4E).

Loop modelling. Two intermediates were calculated at approximately equal maturation distance along the VRC26-UCA to VRC26.01 pathway. Mutations associated with the intermediates were mapped directly onto the structure of VRC26.01. 14 of the 35 residues in the VRC26.01 structure are disordered and were modelled with Loopy[78] (http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Loopy) and represented as grey dots. Mutations of the intermediates were coloured according to approximate time of occurrence based on the longitudinal phylogenetic tree highlighting the timeline of the structural development. These, and the other antibodies with modelled loops (Fig. 4), were modelled in a single loop prediction involving four steps. In the first step, Loopy was used to predict 10 loop conformations. The number of initial loop conformations to be sampled was set to 50,000 (and the not the default value of 2,000). In the second step, all 10 loop conformations were refined using the Protein Preparation Wizard in Maestro (http://www.schrodinger.com/). In the third step, sulphate groups were added to tyrosine at position 100 of the heavy chain and the entire structure was then subjected to all-atom energy minimization in Maestro. A fourth and final step was needed to ensure a reasonable sampling of the rotameric states for the sulphated tyrosines. The Rapid Torsion Scan module in Maestro was used to sample the chi angle involving the sulphate moiety in steps of 20 degrees. The model with the lowest energy after application of the Rapid Torsion Scan module was considered as the best prediction.

Tyrosine sulphation predictions were carried out in GPS-TPS (Z. Pan *et al.*, http://tsp.biocuckoo.org).

Single genome amplification (SGA), sequencing and cloning. HIV-1 RNA was isolated from plasma using the Qiagen QIAamp Viral RNA kit, and reverse transcribed to cDNA using SuperScript III Reverse Transcriptase (Invitrogen, CA). The envelope genes were amplified from single genome templates[49] and amplicons were directly sequenced using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA) and resolved on an ABI 3100 automated genetic analyser. The full-length env sequences were assembled and edited using Sequencher v.4.5 software (Genecodes, Ann Arbor, MI). Multiple sequence alignments were performed using Clustal X (ver. 1.83) and edited with BioEdit (ver. 7.0.9) Sequence alignments were visualized using Highlighter for Amino Acid Sequences v1.1.0 (beta).

For analysis of selection pressure, and to account for recombination between the SU and PI, sequences were partitioned into two alignments (an SU-related, and a PI-related alignment) based on the inferred recombination breakpoints using an in-house script. Breakpoints were identified by a shift in identity from one reference towards the other, and required at least two sequential polymorphisms in common with a corresponding PI/SU-related virus in order to be considered. Phylogenies for both alignments were then reconstructed using FastTree[79] with a GTR+CAT model, and rooted on the PI/SU. Signals of selective pressure were detected with MEME (episodic diversifying selection)[80] and DEPS (directional selection)[81] using the FastTree-generated trees, implemented in Hyphy[82].

The frequencies of specific amino acids at a site and the distribution of net charges in the V2 epitope were calculated from the 2012 filtered web alignment ($n = 3{,}990$) from the Los Alamos HIV database (http://www.hiv.lanl.gov/).

Selected envelope amplicons were cloned into the expression vector pcDNA 3.1 (directional) (Invitrogen) by re-amplification of SGA first-round products using Pfu Ultra II enzyme (Stratagene) with the EnvM primer, 5′-TAGCCCTTCCAGTCCCCCCTTTTCTTTTA-3′ (ref. 83) and directional primer, EnvAstop, 5′-CACCGGCTTAGGCATCTCCTATGGCAGGAAGAA-3′ (ref. 48). Cloned *env* genes were sequenced to confirm that they exactly matched the sequenced amplicon. Autologous clones were mutated at key residues within the C-strand using the Stratagene QuickChange II kit (Stratagene) as described by the manufacturer. Mutations were confirmed by sequencing. Envelope clones were used to generate single round of replication Env-pseudoviruses as described above.

51. van Loggerenberg, F. *et al.* Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS ONE* **3,** e1954 (2008).
52. Doria-Rose, N. *et al.* High throughput HIV-1 microneutralization assay. *Protocol Exchange* http://dx.doi.org/10.1038/protex.2013.069 (2013).
53. Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329,** 856–861 (2010).
54. Julien, J. P. *et al.* Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. *PLoS Pathog.* **9,** e1003342 (2013).
55. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151,** 41–60 (2005).
56. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166,** 95–102 (2009).
57. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166,** 205–213 (2009).
58. Sorzano, C. O. *et al.* A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171,** 197–206 (2010).
59. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).
60. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
61. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
62. Alamyar, E., Giudicelli, V., Li, S., Duroux, P. & Lefranc, M. P. IMGT/HighV-QUEST: the IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* **8** (2012).
63. Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17,** 282–283 (2001).
64. Souto-Carneiro, M. M., Longo, N. S., Russ, D. E., Sun, H. W. & Lipsky, P. E. Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. *J. Immunol.* **172,** 6790–6802 (2004).
65. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731–2739 (2011).
66. Waddell, P. J. & Steel, M. A. General time-reversible distances with unequal rates across sites: mixing gamma and inverse Gaussian distributions with invariant sites. *Mol. Phylogenet. Evol.* **8,** 398–414 (1997).
67. Ashkenazy, H. *et al.* FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40,** W580–W584 (2012).
68. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
69. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
70. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
71. Adams, P. D. *et al.* Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat.* **11,** 53–55 (2004).
72. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32,** W615–W619 (2004).
73. DeLano, W. L. The PyMOL Molecular Graphics System. http://www.pymol.org (DeLano Scientific, San Carlos, California, 2002).
74. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
75. Tran, E. E. *et al.* Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS Pathog.* **8,** e1002797 (2012).
76. Zhou, T. *et al.* Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329,** 811–817 (2010).
77. Pejchal, R. *et al.* A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* **334,** 1097–1103 (2011).
78. Soto, C. S., Fasnacht, M., Zhu, J., Forrest, L. & Honig, B. Loop modeling: Sampling, filtering, and scoring. *Proteins* **70,** 834–843 (2008).
79. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5,** e9490 (2010).
80. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8,** e1002764 (2012).
81. Kosakovsky Pond, S. L., Poon, A. F., Leigh Brown, A. J. & Frost, S. D. A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Mol. Biol. Evol.* **25,** 1809–1824 (2008).
82. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21,** 676–679 (2005).
83. Gao, F. *et al.* The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* **70,** 7013–7029 (1996).

a

```
                                          CDRH1                     CDRH2
IGHV3-30*18    ................................................................................
IGHJ3-01
VRC26-UCA_H    QVQLVESGGGVVQPGRSLRLSCAASGFTFSSYGMHWVRQAPGKGLEWVAVISYDGSNKYYADSVKGRFTISRDNSKNTLY
VRC26-I1       ..............................N.A.D.................S.....N.....................
VRC26-I2       ..............................NFA.G..............F..S.....N.G..................F
VRC26.01       E..V..................T.......NFA.G..............F..S.....N.G.................VF
VRC26.02       E........................Q.S.GN..................A..FA.TKTN...................V.
VRC26.03       E..............K.........R.S.NR..................A.....TD..H..K.W...............
VRC26.04       E..............K.........Q.S.NR..................A.....TD..H..K.W...............
VRC26.05       E........................H.SL...................TG..FA.TKT..G...R....V...I.....F
VRC26.06       E...I.................G....S.NN..................G..FA.IK...GT...............NF.
VRC26.07       E.....................VG.Q.S.NR..................G..F..TDR.H..N.W...............
VRC26.08       E.................T......Q....N..................SV.N..TK..HG...W..............F
VRC26.09       E.................T......Q.N.AN..................S..N..TK..HEE..W......K....S..F
VRC26.10       .AI......................Q....GH.L...............S..FA.TKMD.........A...........
VRC26.11       E...........P..K..T...V..R....AF.................S..FA.IK...........S......~...F
VRC26.12       E....Q..................PY.E.GR..F...............S..FA.RDY.HSP..W...............
```

```
                                      CDRH3
IGHV3-30*18    ..................
IGHJ3-01                                                            ..............
VRC26-UCA_H    LQMNSLRAEDTAVYYCAKDLGESENEEWAT~DYYDFSIGYPGQDPR~~GVVGAFDIWGQGTMVTVSS
VRC26-I1       ......................NK.D....~....L..A..V....~~A..................
VRC26-I2       ............L......V..YKSD....~....L..A..I....~~AM.................
VRC26.01       .......V....L......V.DYKSD..G.~E...I..S..I....~~AM.....L..........P
VRC26.02       ............L.F....IR.Y.C.Y.TS~.....GRPQ.CI.S.~~....T..V...........
VRC26.03       ............L.......R.D.C...WS~.....GKQL.CRKS.~~..A.I..G...........
VRC26.04       ............L.......R.D.C...WS~.....GKQL.CRKS.~~..A.I..K.......I...
VRC26.05       .......PD..M....R.QRYY.C....S~.....GREQ.CL...~~....I..L...........
VRC26.06       .......V....L....R..R.L.C...TLYN....GSRG.CV...~~..A.S..V...........
VRC26.07       ...S........L.......R.D.C...WS~.....GKKL.CRKS.~~..A.V..K...........
VRC26.08       ...S..........F.VR.QR.D.C...WS~.....GREL.CRKF.GL.LA.I.....H........
VRC26.09       .......I....L.F.V..QR.D.C...WS~.....GREL.CRKS.GL.LA.I..M..H........
VRC26.10       .......V....L......MR.Y.C.Y.TS~.....GRPQ.CI.R.~~....I..M..........T
VRC26.11       ....G..G...GL.H.V..MR.L.C....S~.....GKPQ.CL.R.~~..S.ISAW..P........
VRC26.12       .......I....L.F..R..R...C...ES~.....GKKG.CVK..~~..A.GL.L.......I...
```

b

```
                                         CDRL1               CDRL2
IGLV1-51*02    ...............................................................................
IGLJ1-02
VRC26-UCA L    QSVLTQPPSVSAAPGQKVTISCSGSSSNIGNNYVSWYQQLPGTAPKLLIYENNKRPSGIPDRFSGSKSGTSATLGITGLQ
VRC26-I1-L     ..................................................K..E..........................
VRC26-I2-L     ...........................T..........L...........K.D...........................
VRC26.01       ...........................T.........RL...........K.DN......................S...
VRC26.02       .........................N.....KD......I.A.................S..............A.....
VRC26.03       ................................F......R.......I..........ET..............A.....
VRC26.04       .P..............................F......R.....S.....T................A.....A.....
VRC26.05       ..A.....................N~~~...........V..................................V.....
VRC26.06       .........................T.K..Q.S...................DR....T...............A.....
VRC26.07       .A..............N........G......F......R...........S......................A.....
VRC26.08       ..A.....................N.......F......V........L..TYR....................A.....
VRC26.09       ..A.....................N.......F....................R....................A.S...
VRC26.10       ..............................D.......H.............TR....................A.....
VRC26.11       ................N.......PP.H.EK.D.....R...M...MV...SKR..L...V..A.R..S.....V.....
VRC26.12       ...........T....T..V....GT..T...F.....H.................L.V......A........A.....
```

```
                        CDRL3
IGLV1-51*02    ...................T.L
IGLJ1-02                            Y.........
VRC26-UCA L    TGDEADYYCGTWDSSLSAGGVFGTGTKVTVL
VRC26-I1-L     .............T.................
VRC26-I2-L     .............T.................
VRC26.01       .............T...G.............
VRC26.02       .........A...GRV~~~S.I....N.I..
VRC26.03       .A...E...A..SA...SAR......RI...
VRC26.04       .........A..AA..TSAR........I.S
VRC26.05       ......F.....GG..RT...L....R....
VRC26.06       .....E...E...G~~~..............
VRC26.07       .....Y...A..AAR.NSAR......M....
VRC26.08       .........TV.GVRKGV.A...........
VRC26.09       .......W.AV.GVRRG..A...A.......
VRC26.10       .............VRPNR.A...........
VRC26.11       .A...........GRMN~~~~.....T.S..
VRC26.12       ....G....A..SG..N~~~~I......S..
```

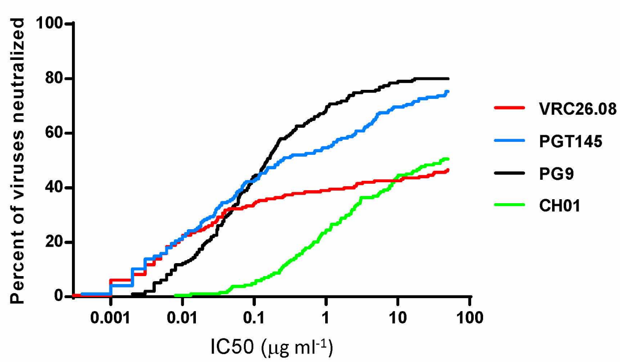
Extended Data Figure 1 | Amino acid sequences of CAP256-VRC26 heavy and light chains. **a**, **b**, Sequences of the 12 B-cell culture derived antibodies, inferred germline *V* and *J* genes, and inferred intermediates are compared to the predicted UCA. **a**, Heavy chain. **b**, Lambda light chain.

a

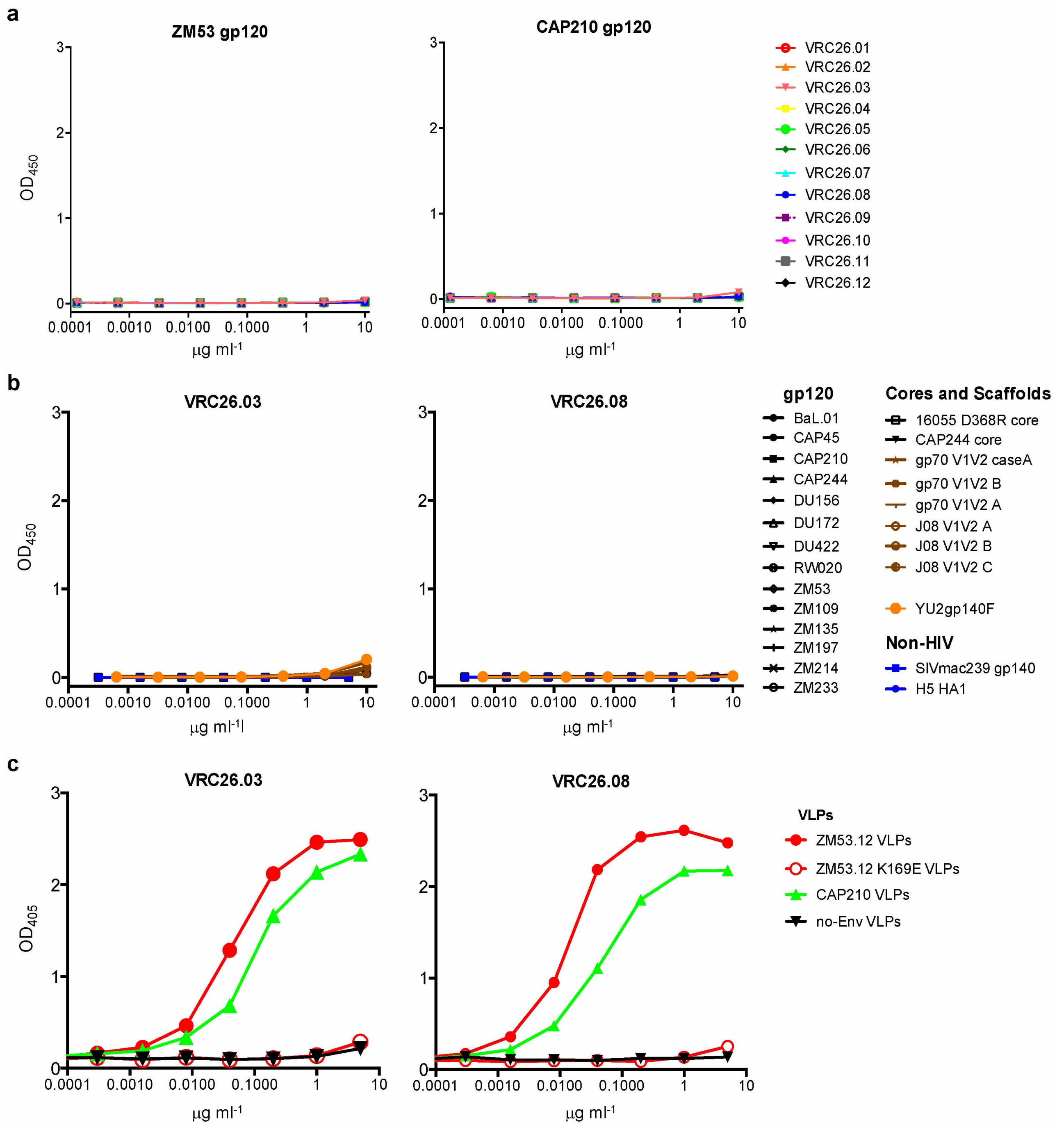
IC50		CAP256-VRC26.01	CAP256-VRC26.02	CAP256-VRC26.03	CAP256-VRC26.04	CAP256-VRC26.05	CAP256-VRC26.06	CAP256-VRC26.07	CAP256-VRC26.08	CAP256-VRC26.09	CAP256-VRC26.10	CAP256-VRC26.11	CAP256-VRC26.12
Autologous	CAP256.PI	>50	>50	>50	>50	>50	0.030	>50	>50	>50	>50	>50	>50
	CAP256.SU	0.14	0.015	0.004	0.003	0.007	0.055	0.008	0.003	0.002	0.031	0.055	0.032

Heterologous by clade	IC50	CAP256-VRC26.01	CAP256-VRC26.02	CAP256-VRC26.03	CAP256-VRC26.04	CAP256-VRC26.05	CAP256-VRC26.06	CAP256-VRC26.07	CAP256-VRC26.08	CAP256-VRC26.09	CAP256-VRC26.10	CAP256-VRC26.11	CAP256-VRC26.12
C	96ZM651.02	>50	>50	>50	>50	>50	>50	>50	2.630	0.991	>50	>50	>50
	CAP210.E8	0.166	0.062	0.008	0.005	0.013	1.200	0.021	0.003	0.002	0.100	0.073	0.072
	CAP244.D3	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	CAP45.G3	31.8	>50	16.0	>50	>50	>50	>50	4.00	16.0	>50	>50	>50
	DU156.12	>50	>50	0.090	10.4	>50	>50	>50	0.006	0.016	>50	>50	>50
	DU172.17	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	DU422.01	>50	>50	0.224	4.17	>50	>50	>50	0.008	0.023	8.20	11.5	>50
	ZM109.4	>50	>50	>50	>50	26.5	>50	>50	>50	>50	>50	>50	>50
	ZM135.10a	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	ZM197.7	22.1	1.97	0.302	0.254	0.113	>50	>50	0.022	0.011	0.346	0.055	>50
	ZM214.15	>50	>50	0.100	>50	>50	>50	>50	3.23	0.333	>50	>50	>50
	ZM233.6	0.701	0.701	0.032	0.060	0.003	22.600	>50	0.001	0.001	0.228	0.322	>50
	ZM249.1	>50	>50	0.004	11.4	>50	>50	44.1	0.076	0.010	>50	34.5	>50
	ZM53.12	0.036	0.043	0.008	0.004	0.012	0.082	0.014	0.004	0.002	0.072	0.081	0.211
A	0260.v5.c36	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	BG505.w6m	>50	>50	4.760	8.430	>50	>50	>50	0.143	0.054	>50	4.200	>50
	KER2008.12	0.308	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	KER2018.11	5.95	0.18	0.014	0.008	0.012	>50	3.100	42.7	0.002	0.335	1.76	>50
	MB201	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	MB539.2B7	>50	>50	>50	>50	>50	>50	>50	31.9	40.7	>50	>50	>50
	Q168.a2	>50	>50	>50	>50	>50	>50	>50	0.148	0.181	>50	>50	>50
	Q23.17	>50	>50	0.867	23.6	>50	>50	>50	0.911	9.50	2.59	>50	>50
	Q259.d2.17	>50	>50	0.002	0.025	0.023	>50	42.5	0.001	0.008	24.6	>50	>50
	Q461.e2	>50	>50	>50	>50	>50	>50	>50	0.502	0.705	>50	>50	>50
	Q769.d22	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	Q842.d12	>50	>50	>50	>50	>50	>50	>50	>50	10.1	>50	>50	>50
	RW020.2	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	UG037.8	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
B	6535.3	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	AC10.29	>50	>50	>50	>50	>50	0.030	>50	>50	>50	>50	>50	>50
	CAAN.A2	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	PVO.04	>50	7.98	>50	>50	44.400	0.947	>50	11.800	>50	2.48	10.8	>50
	QH0692.42	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	REJO.67	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	RHPA.7	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	SC422.8	>50	>50	>50	>50	>50	0.439	>50	>50	>50	>50	>50	>50
	THRO.18	>50	>50	>50	>50	>50	3.30	>50	>50	>50	>50	>50	>50
	TRJO.58	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	TRO.11	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	WITO.33	>50	>50	>50	>50	>50	0.005	>50	>50	>50	>50	>50	>50
D	191821.E6.1	>50	>50	1.50	>50	>50	>50	>50	0.003	0.054	>50	>50	>50
	3016.V5.c36	16.4	0.057	0.450	0.283	0.011	>50	>50	0.005	0.004	0.017	0.300	7.93
	6405.v4.c34	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
AE	C1080.c3	>50	>50	0.006	0.212	>50	>50	>50	0.023	0.019	>50	0.018	>50
	CM244.ec1	3.32	2.030	0.007	0.361	0.517	>50	6.90	1.00	0.105	0.914	26.7	>50
	CNE3	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	TH976.17	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
	% breadth	19%	17%	36%	30%	21%	17%	13%	47%	47%	23%	26%	6%
	geomean IC50	1.88	0.40	0.08	0.32	0.10	0.38	1.51	0.11	0.07	0.60	0.94	0.49

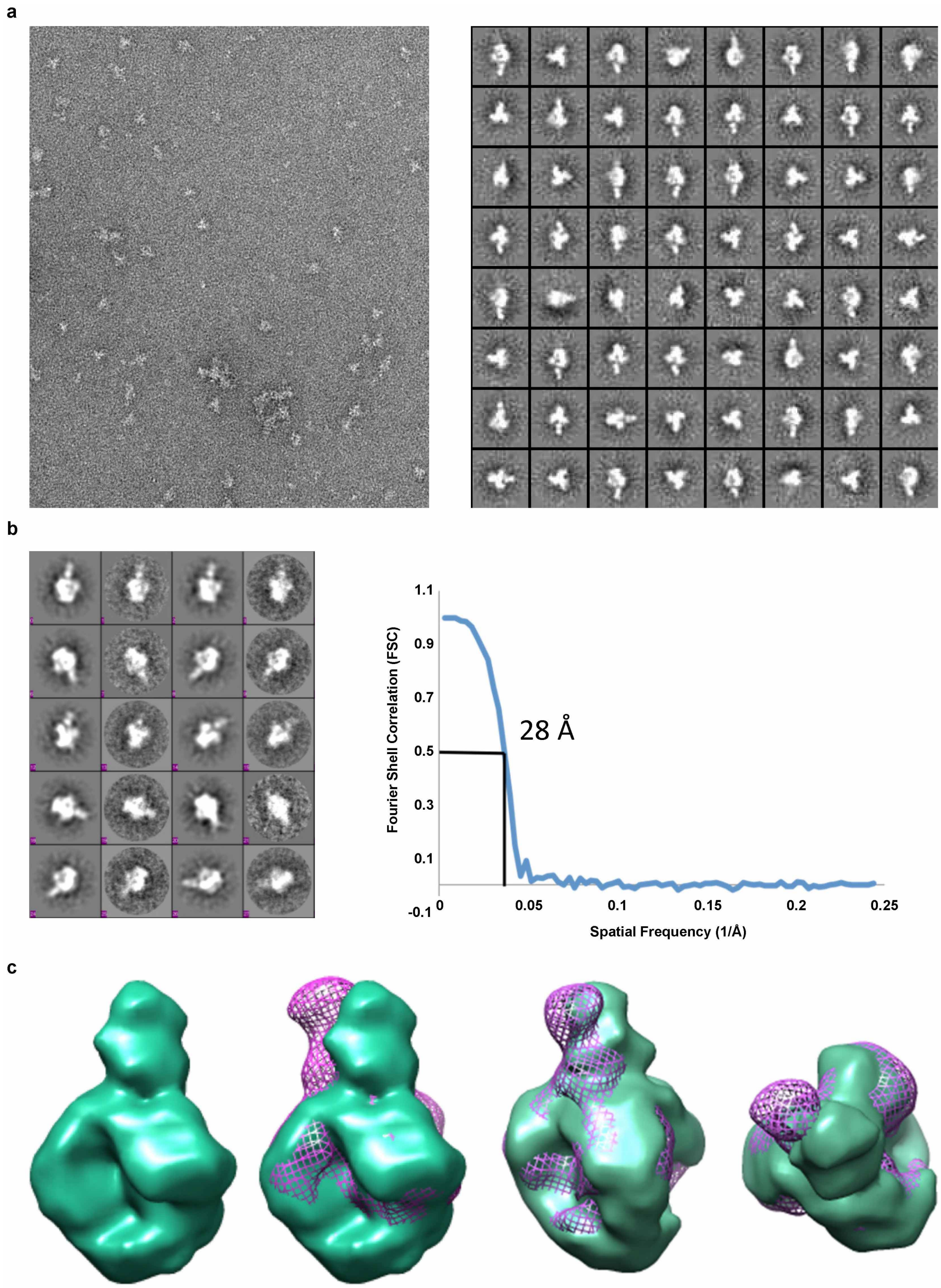
b



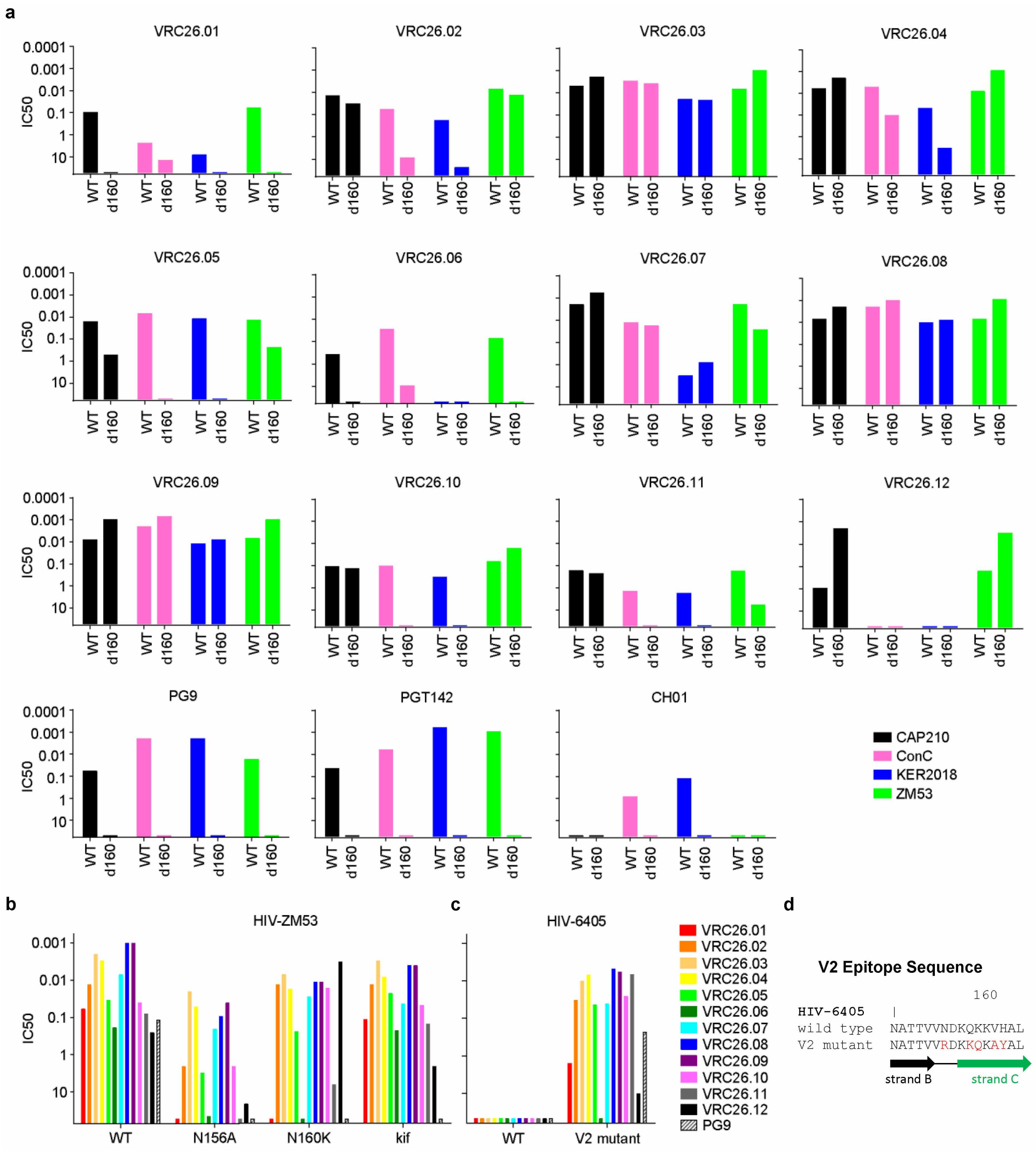
Extended Data Figure 2 | Neutralization breadth and potency of CAP256-VRC26 antibodies. a, Neutralization of autologous (CAP256 PI and SU) and 47 heterologous viruses by CAP256-VRC26 antibodies. Neutralization was measured using a TZM-bl assay with Env-pseudoviruses. Geometric mean was calculated for values $<50\ \mu g\ ml^{-1}$. **b**, Breadth-potency curves. Neutralization of a 194-virus panel was measured for VRC26.08, PG9, PGT145 and CH01. The curves show the percent of viruses neutralized at any given IC_{50}.



Extended Data Figure 3 | CAP256-VRC26 antibodies recognize a quaternary epitope. **a**, All 12 CAP256-VRC26 monoclonal antibodies were tested by ELISA against gp120 from ZM53 and CAP210. Positive control antibody PG9 bound to both gp120s (not shown). **b**, Twenty-three proteins and scaffolded V1V2 constructs were tested by ELISA for binding of CAP256-VRC26.03 and CAP256-VRC26.08. PG9 bound to several of these (not shown). Similar data were observed for CAP256-VRC26.06, .07 and .09. **c**, Binding of CAP256-VRC26.03 and CAP256-VRC26.08 to virus-like particles (VLP). VLP expressing ZM53, ZM53.K169E, CAP210 or no Env were concentrated by pelleting and used to coat ELISA plates; assays were performed without detergent to preserve the trimer spikes. Similar data were observed for CAP256-VRC26.06, .07 and .09.



Extended Data Figure 4 | Visualization of CAP256-VRC26.09 bound to Env trimers by negative-stain electron microscopy. **a**, Raw micrograph and corresponding reference free 2D class averages of VRC26.09 in complex with cleaved soluble BG505 SOSIP.664 gp140 trimers. **b**, Projection matching of 3D model refinement and FSC curve used to calculate resolution. Resolution, 28Å at FSC = 0.5. **c**, 3D reconstruction of VRC26.09:BG505 SOSIP.664 complex (green surface) alone and overlayed with PG9:SOSIP (purple mesh). The reconstructions are nearly identical in the trimer portion while displaying small differences in the Fab angles.



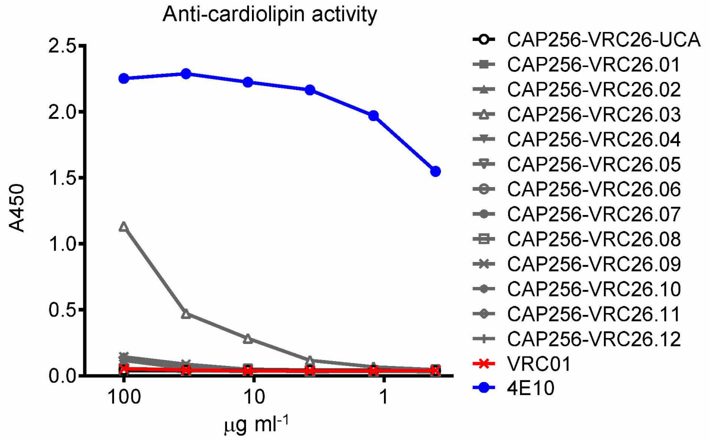
Extended Data Figure 5 | Effects of V2 mutations on neutralization activity of CAP256-VRC26 antibodies. **a**, Each panel shows neutralization of wild-type and N160 glycan mutant CAP210.E8, ConC, KER2018.11 and ZM53.12 viruses. CAP256-VRC26 monoclonal antibodies are partially and variably affected by loss of N160 glycan, in a virus-strain specific manner. In contrast, PG9-class antibodies PG9, PGT142, and CH01 are uniformly knocked out by N160 mutation. **b**, CAP256-VRC26 monoclonal antibodies are partially and variably affected by changes in V2 glycans. Neutralization by each antibody was measured against wild-type ZM32.12, mutants N156A and N160K, and ZM53.12 grown in the presence of kifunensine, an inhibitor of glycan processing. In contrast to CAP256-VRC26 antibodies, PG9 activity is knocked out by the mutations and by kifunensine. **c**, HIV-6405 wild type is resistant to PG9 and CAP256-VRC26 antibodies, and its PG9-sensitive mutant[34] is also sensitive to CAP256-VRC26 antibodies. **d**, Sequences of wild type and mutant HIV-6405.

a

```
                                                         CDR H1                    CDR H2
UCA               QVQLVESGGG VVQPGRSLRL SCAASGFTFS SYGMHWVRQA PGKGLEWVAV ISYDGSNKYY ADSVKGRFTI
038-190676 (1x)   .......... .......... .......... .......... .......... .......... ..........
038-195559 (1x)   .......... .....G.... .......S.. .......... .......... .......... ..........
038-006719 (35x)  .......... .......... .......S.. .......... .......... .......... ..........
038-223402 (1x)   .......... .......... .......S.. .......... .......... .......... ..........
038-226205 (1x)   .......... .I........ .......S.. .......... .......... .......... ..........

                                                              CDR H3
UCA          SRDNSKNTLY LQMNSLRAED TAVYYCAKDL GESENEEWAT DYYDFSIGYP GQDPRGVVGA FDIWGQGTMV TVSS
038-190676   .......... .......... .......... .........S .....R.... .......... .......... ....
038-195559   .......... .......... ..L....... .........S .....GRD.. .........T .......... ....
038-223402   .......... .......... ..L....... .........S .....GRD.. .........T .......... ....
038-223402   .......... .......... ..L-...... .........S .....GRD.. .........T .......... ....
038-226205   .......... .......... ..L....... .........S .....GRD.. .........T .......... ....
```

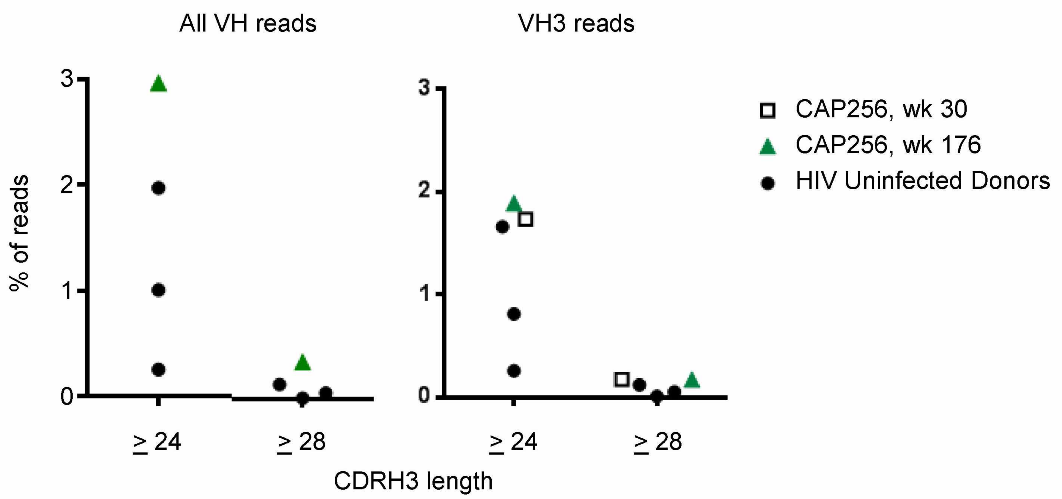
b



c

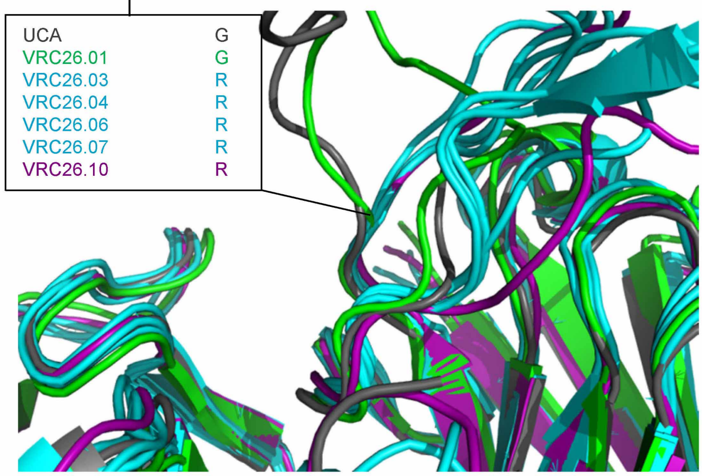
Antibody	Hep2 staining
CAP256-VRC26-UCA	-
CAP256-VRC26.01	-
CAP256-VRC26.02	-
CAP256-VRC26.03	-
CAP256-VRC26.04	-
CAP256-VRC26.05	-
CAP256-VRC26.06	-
CAP256-VRC26.07	-
CAP256-VRC26.08	-
CAP256-VRC26.09	-
CAP256-VRC26.10	-
CAP256-VRC26.11	-
CAP256-VRC26.12	-
VRC01	-
4E10	+

d



Extended Data Figure 6 | Origins of long CDR H3s in donor CAP256. **a**, Week 38 sequences from 454 that support the calculation of the UCA. Unique amino acid sequences with 2–5 residue changes in the CDR H3 are compared to the calculated UCA sequence. Each contained fewer than 3 combined nucleotide mutations in VH and JH. Parentheses, number of corresponding reads in the raw 454 data. **b**, **c**, Lack of autoreactivity. **b**, ELISA for binding to cardiolipin. 4E10 was strongly positive, CAP256-VRC26.03 was weakly positive, and the other 11 CAP256-VRC26 monoclonal antibodies and the UCA were negative along with control antibody VRC01. **c**, Staining on Hep2 cells was assessed at 50 and 25 μg ml^{-1}. Only the positive control, mAb 4E10, showed positive staining. **d**, Distribution of CDRH3 lengths among 454 sequencing reads of B cell transcripts. The percentage of high-quality NGS reads that have CDR H3 ≥ 24 or ≥ 28 are shown for three HIV-1 uninfected donors (solid circles on both right and left plots) and for donor CAP256 (week 176) amplified with all-VH primers donor, and CAP256 (week 30) amplified with VH3 primers. High-quality reads are defined as successful V and J assignments and a continuous open reading frame. CDRH3 lengths use the IMGT definitions.

a



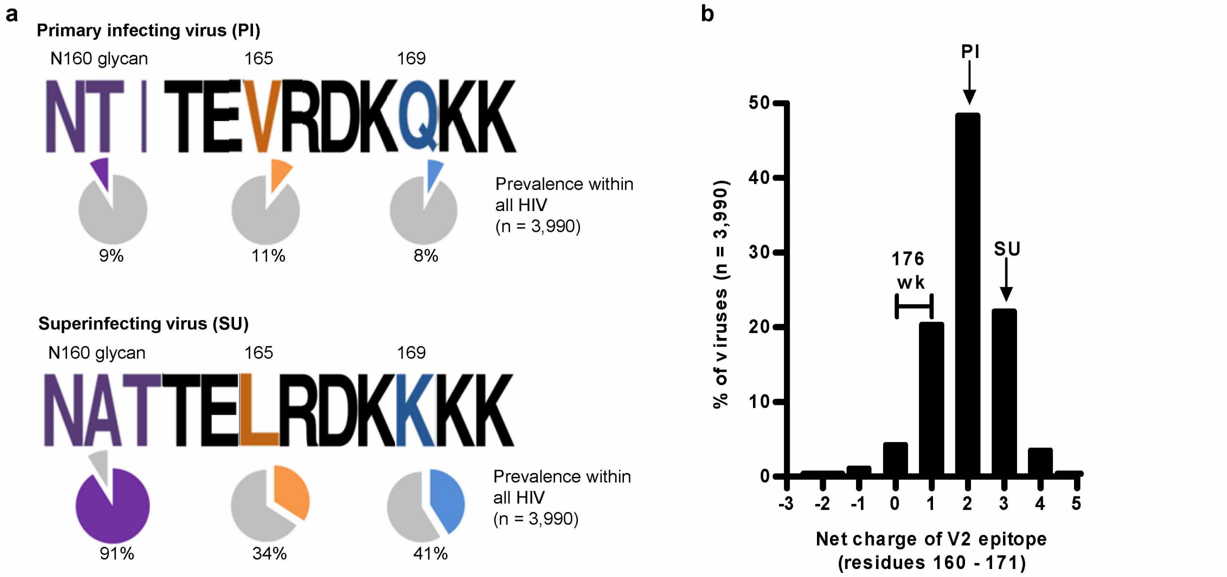
c

	VRC26.03	
IC50	wild type	- disulphide
CAP256-PI	>50	>50
CAP256-SU	0.004	3.1
3016.v5.v45	0.118	>50
CAP210.E8	0.011	>50
CM244.ec1	0.000	>50
KER2008.12	>50	>50
KER2018.11	0.003	>50
Q259.17	0.014	>50
ZM197.7	0.160	>50
ZM233.6	0.032	>50
ZM53.12	0.008	>50

b

```
VRC26.01          CAKDVGDYKSDEWGTEYYDISISYPIQDPRAMVGAFDLW
VRC26.03          CAKDLREDECEEWWSDYYDFGKQLPCRKSRGVAGIFDGW
VRC26.03_mutant   CAKDLREDESEEWWSDYYDFGKQLPIRKSRGVAGIFDGW
```

Extended Data Figure 7 | Loss of flexibility at the base of the CDR H3. **a**, Top shows logograms of CDR H3 sequences extracted from the heavy chain phylogenetic tree from weeks 59 and 119. The height of each letter is proportional its frequency in the population. Sequences that lack a disulphide bond contain a highly conserved glycine at the third position of the CDR H3 (residue 97, Kabat definition). The appearance of the two cysteines that form the disulphide bond coincides with a glycine to arginine mutation at this site. Bottom shows overlay and close-up of crystal structures from Supplementary Fig. 6A. Loss of the glycine limits flexibility at the base of the CDR H3 and is shown in the crystal structures to be the initial site of divergence in the CDR H3 loops between the antibodies without the disulphide bond (UCA and CAP256-VRC26.01) and those with it (CAP256-VRC26.03, .04, .06, .07, .10). This mutation may contribute to the conserved trajectory of the CDR H3 protrusion towards the heavy chain that is seen in the more mature antibody structures. **b**, CDRH3 and flanking sequences for VRC26.01, VRC26.03, and a mutant VRC26.03 in which the conserved cysteines are changed to the corresponding amino acids found in VRC26.01. **c**, Neutralization activity of VRC26.03 and the mutant shown in panel **b**. The mutant shows reduced activity against CAP256 SU and complete loss of heterologous activity.



c Infecting viruses and effect of polymorphisms in V2 epitope

Env clone	Residues 160 - 171	Neutralization IC$_{50}$ titers (μg ml^{-1})											
PI	NTITEVRDKQKK	>50	>50	>50	>50	>50	0.03	>50	>50	>50	>50	>50	>50
SU	NATTELRDKKKK	0.04	0.008	0.003	0.005	0.009	0.06	0.008	0.004	0.003	0.01	0.03	0.01
SU T162I	NAITELRDKKKK	0.02	0.05	0.002	0.003	1.7	0.04	0.01	0.005	0.004	0.4	0.1	0.1
SU L165V	NATTEVRDKKKK	0.05	0.01	0.003	0.004	0.007	0.05	0.04	0.006	0.008	0.02	0.04	0.05
SU K169Q	NATTELRDKQKK	>50	0.05	0.03	0.06	0.1	0.8	1.1	0.04	0.02	0.10	0.09	>50
		01	02	03	04	05	06	07	08	09	10	11	12

d Validation of escape mutations

Env clone	Residues 160 - 171	Neutralization IC$_{50}$ titers (μg ml^{-1})												Mutations first detected
SU K169I	NATTELRDKIKK	>50	0.1	0.02	0.05	0.20	0.05	1.4	0.04	0.02	0.08	0.1	>50	34 weeks
SU K171N	NATTELRDKKKN	0.1	0.03	0.004	0.01	0.03	0.16	0.02	0.005	0.004	0.03	0.02	>50	94 weeks
SU R166K	NATTELKDKKKK	>50	>50	4.8	16	>50	>50	>50	0.1	0.2	>50	>50	>50	
SU R166S	NATTELSDKKKK	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	176 weeks
SU K169E	NATTELRDKEKK	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	
		01	02	03	04	05	06	07	08	09	10	11	12	

e Reversion of I169 escape mutation in a 34 week clone

34 wk clones	Residues 160 - 171	Neutralization IC$_{50}$ titers (μg ml^{-1})											
34 wk wt	NATTELRDKIKK	>50	0.05	0.02	0.7	0.1	0.2	1	0.05	0.03	0.08	0.04	>50
34 wk I169K	NATTELRDKKKK	0.05	<0.02	<0.02	<0.02	<0.02	<0.02	<0.02	<0.02	<0.02	<0.02	<0.02	<0.02
		CAP256-VRC26.01	CAP256-VRC26.02	CAP256-VRC26.03	CAP256-VRC26.04	CAP256-VRC26.05	CAP256-VRC26.06	CAP256-VRC26.07	CAP256-VRC26.08	CAP256-VRC26.09	CAP256-VRC26.10	CAP256-VRC26.11	CAP256-VRC26.12

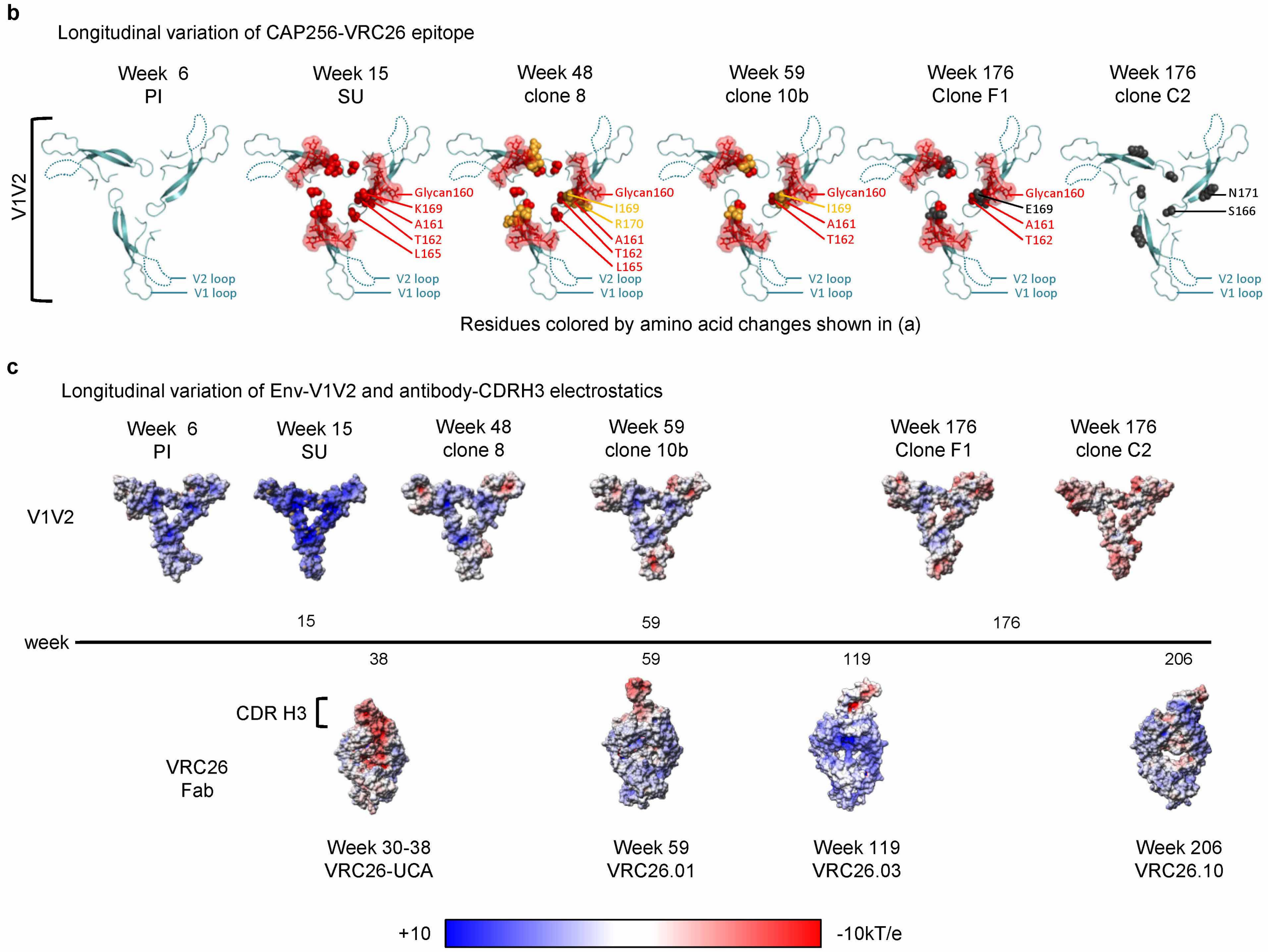
Extended Data Figure 8 | Viral polymorphisms and escape mutations. **a**, Frequency of CAP256 PI and SU polymorphisms at positions 160–162 (glycosylation sequon), 165 and 169. Coloured slices on pie charts and percentages indicate prevalence of these polymorphisms within global circulating viruses in the Los Alamos Sequence Database (*n* = 3,990). **b**, Distribution of net charge of the V2 epitope, defined as residues 160–171, within global circulating viruses (*n* = 3,990). The charge of the PI, SU and 176 week clones are indicated. **c**, CAP256-VRC26 monoclonal antibody neutralization of the SU and PI viruses, and of the SU virus mutated to contain PI polymorphisms 162I, 165V or 169Q. **d**, CAP256-VRC26 monoclonal antibody neutralization of the SU virus mutated to contain known CAP256 escape mutations in the V2 epitope. **e**. CAP256-VRC26 monoclonal antibody neutralization of 34 week clone (designated wild type, wt) with an SU-like V1V2, compared to the I169K back mutant. **c–e,** The V2 epitope sequence, with mutated residues in red is shown on the left, IC$_{50}$ values in the middle, and the time point when mutations were first detected in Env sequences on the right (weeks post-infection).

a Longitudinal variation of Env and sensitivity to CAP256-VRC26

Env clones	Residues 160–171	UCA	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.11	.12	Charge (monomer)
PI	NTITEVRDKQKK	>50	>50	>50	>50	>50	>50	0.03	>50	>50	>50	>50	>50	>50	+2
SU	NATTELRDKKKK	>50	0.04	0.008	0.003	0.005	0.009	0.06	0.008	0.004	0.003	0.01	0.01	0.01	+3
48 week clone 8	NATTELRDKIRK	>50	>50	0.2	<0.02	0.03	0.2	0.25	0.5	0.040	<0.02	0.1	0.5	>50	+2
59 week clone 10b	NATTEVRDKIKK	>50	>50	0.35	0.19	0.26	0.72	0.20	5.43	0.05	0.04	0.27	0.54	>50	+2
176 week clone F1	NATTEVRDKEKK	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	+1
176 week clone C2	NTITEVSDKQKN	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	0

CAP256-VRC26 neutralization (IC_{50} μg ml^{-1})

<0.1 | 0.1-1 | 1-49 | >50 (IC_{50} μg/ml)



Extended Data Figure 9 | Longitudinal changes in CAP256 V1V2. **a**, Variation in the V1V2 sequence of six Env clones. Amino acid mutations from residues 160–171 are highlighted and corresponding changes in neutralization for the six Env clones by CAP256-VRC26.01-.12 and the UCA are shown. The charge of the displayed sequences that make up the central region of the trimer are shown on the right. **b**, Residue changes highlighted in **a** were mapped onto the V1V2 domain in the crystal structure of the HIV-1 BG505.664 SOSIP Env trimer. The structure is viewed looking towards the viral membrane along the trimer axis. Mutations are coloured as in panel **a** and represented as spheres (amino acids) or stick and surface (glycan). **c**, Electrostatic surface representations of the full V1V2 region for each Env clone(top row), Fabs (bottom row). Timeline of infection is shown in the middle. V1V2 sequences were modelled with SWISS-MODEL using the BG505.664 SOSIP as a template. Escape mutations R166S, K171N and K169E resulted in a net charge change in the V2 epitope from +3 (SU) to a rare 0. Antibody CDR H3s became less negatively charged over time, suggesting co-evolution of the viral epitope and the antibody paratope.

Extended Data Table 1 | Genetic characteristics of CAP256-VRC26 antibodies and V1V2-directed broadly neutralizing antibodies from other donors

a

Donor	Antibody	VH gene	VL gene	% Divergence, Nucleotide from VH	from VL	from UCA-H	from UCA-H CDRH3
CAP256	*VRC26-I1*	IGHV3-30*18	IGLV1-51*02	2%	4%	3%	8%
	VRC26-I2			6%	7%	7%	13%
	VRC26.01			8%	4%	11%	18%
	VRC26.02			9%	5%	13%	27%
	VRC26.03			9%	7%	13%	28%
	VRC26.04			9%	8%	13%	28%
	VRC26.05			10%	5%	13%	24%
	VRC26.06			11%	7%	16%	30%
	VRC26.07			12%	8%	14%	26%
	VRC26.08			12%	10%	17%	33%
	VRC26.09			14%	10%	19%	34%
	VRC26.10			12%	4%	15%	26%
	VRC26.11			12%	14%	18%	32%
	VRC26.12			15%	8%	18%	27%
IAVI24	PG9	IGHV3-33*05	IGLV2-14*01	12%	8%		
	PG16			15%	12%		
CH0219	CH01	IGHV3-20*01	IGKV3-20*01	16%	11%		
	CH02			15%	14%		
	CH03			14%	11%		
	CH04			14%	11%		
IAVI84	PGT141	IGHV1-8*01	IGKV2-28*01	16%	13%		
	PGT142			16%	13%		
	PGT143			16%	13%		
	PGT144			17%	12%		
	PGT145			17%	17%		

b

Donor	Antibody	VH gene	VL gene	CDRH3 Length	% Divergence, Amino Acid from VH	from VL	from UCA-H	from UCA-H CDRH3
CAP256	*VRC26-I1*	IGHV3-30*18	IGLV1-51*02	35	6%	5%	8%	20%
	VRC26-I2			35	11%	7%	13%	29%
	VRC26.01			35	16%	7%	18%	40%
	VRC26.02			35	16%	7%	27%	46%
	VRC26.03			35	14%	9%	28%	46%
	VRC26.04			35	14%	8%	28%	46%
	VRC26.05			35	20%	8%	24%	40%
	VRC26.06			36	18%	9%	30%	44%
	VRC26.07			35	18%	9%	26%	46%
	VRC26.08			37	16%	9%	33%	51%
	VRC26.09			37	21%	7%	34%	54%
	VRC26.10			35	17%	5%	26%	46%
	VRC26.11			35	23%	23%	32%	49%
	VRC26.12			35	22%	14%	27%	43%
IAVI24	PG9	IGHV3-33*05	IGLV2-14*01	28	20%	15%		
	PG16			28	21%	21%		
CH0219	CH01	IGHV3-20*01	IGKV3-20*01	24	29%	17%		
	CH02			24	22%	23%		
	CH03			24	22%	19%		
	CH04			24	23%	17%		
IAVI84	PGT141	IGHV1-8*01	IGKV2-28*01	32	28%	21%		
	PGT142			32	30%	21%		
	PGT143			32	28%	22%		
	PGT144			32	31%	23%		
	PGT145			31	30%	26%		

a, **b**, Data are from the present study and from references 19–21. CAP256-VRC26.01-12 are derived from B cell culture, while CAP256.VRC26-I1 and –I2 (in italics) are inferred intermediates. CDRH3 lengths use Kabat notation. **a**, Nucleotides. **b**, Amino acids.

doi:10.1038/nature13191

Molecular photons interfaced with alkali atoms

Petr Siyushev[1], Guilherme Stein[1], Jörg Wrachtrup[1,2] & Ilja Gerhardt[1,2]

Future quantum communication will rely on the integration of single-photon sources, quantum memories and systems with strong single-photon nonlinearities[1]. Two key parameters are crucial for the single-photon source: a high photon flux with a very small bandwidth, and a spectral match to other components of the system. Atoms or ions may act as single-photon sources—owing to their narrowband emission and their intrinsic spectral match to other atomic systems—and can serve as quantum nonlinear elements. Unfortunately, their emission rates are still limited, even for highly efficient cavity designs[2]. Single solid-state emitters such as single organic dye molecules are significantly brighter[3] and allow for narrowband photons[4]; they have shown potential in a variety of quantum optical experiments[5,6] but have yet to be interfaced with other components such as stationary memory qubits. Here we describe the optical interaction between Fourier-limited photons from a single organic molecule and atomic alkali vapours, which can constitute an efficient quantum memory. Single-photon emission rates reach up to several hundred thousand counts per second and show a high spectral brightness of 30,000 detectable photons per second per megahertz of bandwidth. The molecular emission is robust and we demonstrate perfect tuning to the spectral transitions of the sodium D line and efficient filtering, even for emitters at ambient conditions. In addition, we achieve storage of molecular photons originating from a single dibenzanthanthrene molecule in atomic sodium vapour. Given the large set of molecular emission lines matching to atomic transitions, our results enable the combination of almost ideal single-photon sources with various atomic vapours, such that experiments with giant single-photon nonlinearities, mediated, for example, by Rydberg atoms[7,8], become feasible.

The emission wavelength of the organic dye dibenzanthanthrene (DBATT, $C_{30}H_{16}$, Fig. 1a, inset) in a *n*-tetradecane Shpol'skiĭ matrix[9] matches well with the sodium-D-line transitions around 589 nm. Under cryogenic conditions, photons from the lowest electronic transition are usually transform-limited with a spectral width of around 17 MHz (excited-state lifetime, 9.4 ± 0.5 ns)[9,10]. Figure 1a displays the excitation spectra of several thousand single DBATT molecules. Usually, the inhomogeneous spread of the molecular transition is considered disadvantageous, but it is an advantage in the experiments presented here, because molecular emissions in resonance with the sodium-D-line transitions can be found. To confirm the spectral match between the molecules and sodium atoms, we perform simultaneous spectroscopy of molecules and atomic vapour around the sodium D_2 line (589.1583264 nm, in vacuum). Figure 1b shows a more finely resolved recording of DBATT molecules in our sample. Parallel to this, Doppler free spectroscopy on sodium vapour is performed. A small magnetic field (2 mT) spectrally splits the atomic absorption lines. These are later subtracted to obtain a dispersive signal (Fig. 1b, green). This technique is known as dichroic atomic vapour laser lock[11]. A similar recording of atoms and single molecules was obtained for the D_1 line (not shown). Almost all of the investigated molecules exhibit high brightness (more than 10^4 counts per second (c.p.s.) at an excitation of 10 nW) and undetectable spectral diffusion. As an example, a time trace recorded over a minute is depicted in Fig. 1c. In all our studies the molecules were chosen at the beginning of the experiment and studied for several hours. This shows that DBATT in tetradecane has a usable overlap with the sodium transitions, without noise introduced by spectral diffusion.

With molecules matching the sodium transition, we use spectroscopy to combine the two techniques. The optical density of a resonant vapour cell reaches very high values, because the atomic density—and, therefore, the optical absorption—displays an approximately exponential behaviour with increasing temperature. To explore the full filtering capability of the sodium cell, we replaced the commercial long pass filter (Semrock, 594LP), used in the above experiments with a hot sodium vapour cell. A recorded excitation spectrum is shown in Fig. 2a. Single molecules are clearly identified on a base line representing the dark

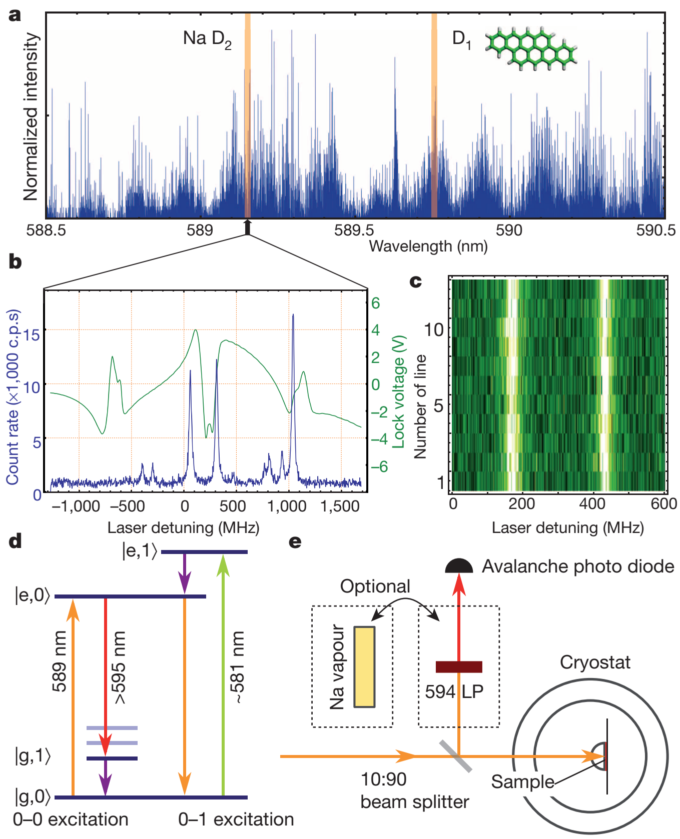


Figure 1 | Spectral match of DBATT to atomic sodium. a, Excitation scan, showing the inhomogeneous band of DBATT molecules across the sodium-D-line transitions. **b**, Simultaneous recorded spectra for single molecules (blue) and spectroscopy of the sodium D_2 transition (green). **c**, A time trace of 12 lines, recorded in a run of 1 min. **d**, Simplified level scheme for the selected molecule. $|g, N\rangle$, electronic ground state, $|e, N\rangle$ electronic excited state, with an Nth vibrational excitation. **e**, The experimental setup is a cryogenic confocal microscope ($T = 1.4$ K), which uses a solid immersion lens to enhance the collection efficiency from the selected molecule. In the detection arm, equipped with a single-photon detector (avalanche photo diode), either a commercial long-pass filter (594LP) or a hot atomic vapour cell can be placed (Na, sodium).

[1]3. Physikalisches Institut, Universität Stuttgart, Stuttgart Research Center of Photonic Engineering (SCoPE), and the Center for Integrated Quantum Science and Technology (IQST), Pfaffenwaldring 57, 70569 Stuttgart, Germany. [2]Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany.

count level of the used single-photon detector. The cell temperature was $T = 155\ ^{\circ}\mathrm{C}$. This results in an optical rejection (reduction of the incident light) for the excitation laser of five orders of magnitude, whereas all emitted redshifted photons can pass through the cell. Off-band, a fraction of a nanometre detuned from the sodium line, the transmission is approximately 85%, limited solely by the cell windows. For these experiments, the laser excitation wavelength has to be spectrally within the Doppler broadening of the atomic vapour. Laser detuning shows, when scanning over the edge of the sodium filter, that the detection is dominated by the back-scattered laser, which is orders of magnitude stronger than the single-molecule signals (Figure 2a). The net transmission through the filter is higher than for commercial dichroic filters, owing to the extremely narrow block-band (spectral rejection). This allows for more detectable photons.

To identify molecules that exactly match one defined wavelength, we locked the laser on the crossover resonance of the sodium D_2 line (falling slope, midway between the $F = 1$ and $F = 2$ signals, see Fig. 2a), and performed confocal scanning over the sample area. This is a rare case: a cryogenic single-molecule study performed with an absolute frequency measure, referenced by an atomic transition. Numerous molecules were found, and efficiently excited with saturation intensities in the few-microwatt range. To compare the filtering capabilities to those of a commercial filter, we perform lateral scanning in both configurations. The commercial filter represents a good choice for this wavelength and displays a pass-band (transmission of the incident light) of more than 97%. Figure 2b shows a lateral scan with the commercial filter, whereas Fig. 2c shows the same area with filtering based solely on sodium vapour. The commercial filter yields a maximum detectable count rate of 8×10^4 c.p.s. The higher transmission resulting from the smaller block-band means that the sodium filter allows a maximum count rate of 1.4×10^5 c.p.s. with the same molecule. The achievable signal-to-noise ratio with 2 ms of integration time per pixel is measured to be 53 with the commercial filter and 68 while filtering with the sodium cell.

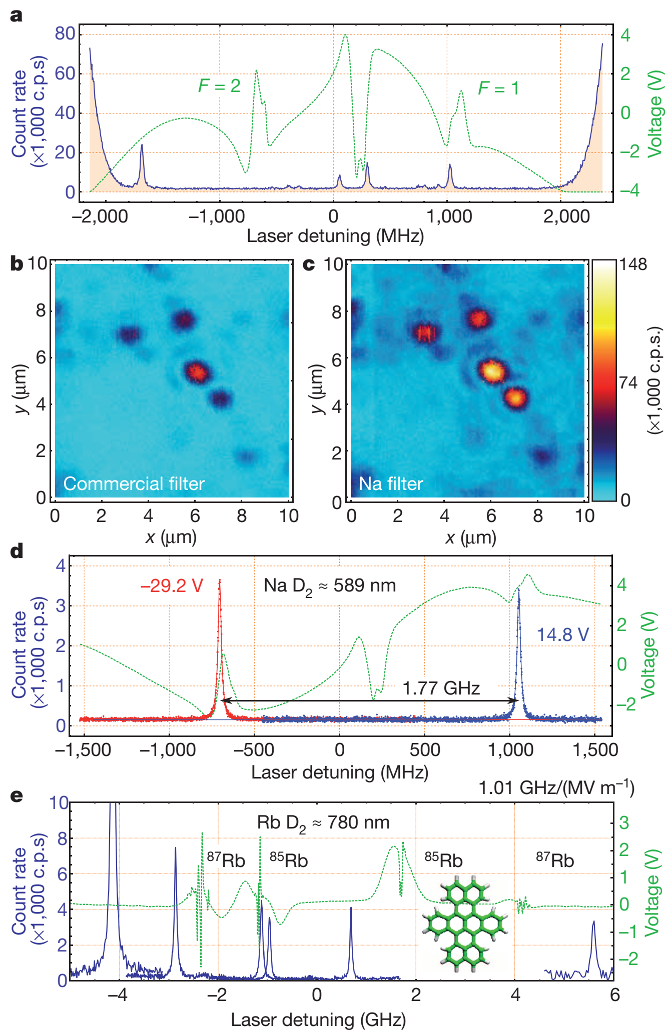


Figure 2 | Spectroscopy and microscopy of molecules and atoms. **a**, Doppler free spectrum (dichroic atomic vapour laser lock) of the sodium D_2 line (green), with simultaneously recorded excitation spectrum of single molecules (blue). The detection is solely realized through an atomic filter. **b**, Recorded lateral scan across the sample with a commercial long-pass filter (594LP). **c**, The same area, filtering the emission solely with a hot sodium cell. The overall signal rate is increased, and the signal-to-noise ratio is enhanced. **d**, Stark tuning of a single molecule to the hyperfine transitions of the sodium D_2 line. With an electric field of $-1.17\ \mathrm{MV\,m^{-1}}$, the $F = 2$ ground state is reached; with a field of $0.59\ \mathrm{MV\,m^{-1}}$, the $F = 1$ ground state is reached. Effectively, this results in a Stark tunability of $1.01\ \mathrm{GHz/(MV\,m^{-1})}$. **e**, Such experiments can also be performed on the Rb transitions in conjunction with the DBT molecule ($C_{38}H_{20}$, structure in inset to **e**).

We tested the atomic filtering against other broad-band emitters at ambient conditions (fluorescent single-molecule labels and fluorescent beads; see Extended Data Fig. 1), and found an enhancement in the collection efficiency. This is remarkable because the commercial filter has a pass-band with more than 97% transmission. In this study, the vapour cell used was not anti-reflection-coated and the maximum transmission was limited by reflection at the air–glass interface. With better-engineered vapour cells, the detectable number of photons could be improved further. The D_2 transitions are preferred over the D_1 lines because their absorption oscillator strength f is higher by a factor of two.

Most molecules do not emit exactly in resonance with atomic sodium. To solve this problem, we introduce Stark tuning of single molecules[12,13]. By using interdigitated gold electrodes (25 μm spacing), the resonance of one selected molecule under study can be shifted (Fig. 2d). The transition is moved across the entire sodium D_2 line. The spectral response of the molecules is approximately linear versus voltage, and all desired spectral positions of the sodium spectrum can be matched. The molecule DBATT (matched to atomic sodium) is not the only molecule with a matching atomic transition. We evaluated whether other molecular species generate a high flux of single photons at various atomic transition wavelengths. Table 1 lists a number of candidates.

For the often explored alkalis—potassium (K) and rubidium (Rb)—we performed experiments in conjunction with the molecule dibenzoterrylene (DBT) embedded in naphthalene. As before, atomic spectra were recorded along with the molecular transitions. The spectral linewidth was measured between 45 MHz and 55 MHz, which again represents an excellent overlap with the atomic transitions of K and Rb. Figure 2e presents the match between the rubidium D_2 line and DBT in naphthalene. Furthermore, both D-line transitions of atomic potassium—D_2 at 766 nm (see Extended Data Fig. 2) and D_1 at 770 nm (not shown)—are matched. Experiments on the rubidium D_1 line (795 nm) can be performed with DBT in anthracene[14], which is reported to have a spectral insertion site centred around 794.3 nm. For Rydberg excitations into higher states, perylene and its derivatives are reported to generate blue-light single photons between 440 nm and 450 nm (ref. 15). As an example, this would allow for an excitation of sodium from $3^2P_{3/2}$ to $9^2S_{1/2}$.

All of the experiments above were performed in fluorescence excitation, that is, the laser was tuned to the atomic resonances and the emitted redshifted photons were detected. To achieve a single-photon emission, matched to the atomic line, we excited the system into the first vibrational level of the electronic excited state[16,17]. The following experiments can be performed with all the molecules mentioned above and in Table 1, but, given its superior stability, we chose to use DBATT interfaced with sodium. First, a molecule with small detuning to the sodium transition was identified (0–0 excitation, Fig. 1d). Second, the laser was blue-detuned to $\lambda \approx 580.8$ nm (0–1 excitation, Fig. 1d). As a result, the single molecule emits a large number of photons at the 0–0 transition, matching the sodium D line. To reduce the background, which originates from the vibrational decay of the molecule, we introduce a 0.5-nm-wide filter

Table 1 | Candidates for atom–molecule interplay

Atomic species	Wavelength (nm)	Single molecule	Approximate linewidth (MHz)	Reference
Lithium	671.0	TDI	65	26, 27
Sodium	589.0, 589.6	DBATT	20	9, 10, and this study
Potassium	766, 770	DBT	50	14, 28, and this study
Rubidium	780, 795	DBT	50	14, 28, and this study
Helium	587.6	DBATT	20	9, 10
Hydrogen	656.3	TDI	65	26, 27
Mercury	576.9, 579.1	Terrylene	30	29, 30
Sodium-Rydberg	442	Perylene	50	15

TDI, terrylenediimide; DBATT, dibenzanthanthrene; DBT, dibenzoterrylene.

with 42% peak transmission (Omega Optical), which filters out the narrowband zero-phonon emission. To block out scattered light further, a 585-nm long-pass filter (Omega Optical) is used in the detection path.

A convenient method of spectrally selecting the molecular emission that matches the atomic transitions is the use of atomic vapour and the configuration in a Faraday anomalous dispersion optical filter (FADOF)[18,19]. A cell of hot atomic vapour is placed between two crossed polarizers and a longitudinal magnetic field is applied. Because the electric susceptibility for the optical σ^+ and σ^- components differs, a net rotation for the polarized light while passing through the cell can be achieved. When the Doppler-broadened absorption is weak and the net optical rotation reaches 90°, the filter allows transmission through the second polarizer. A scheme of the filter is displayed in Fig. 3a. By varying the magnetic field and temperature, it is possible to spectrally select molecules around the Doppler edge of the sodium transition (Fig. 3b). The transmission of the filter for the selected molecule is above 80%. In this configuration a selected molecule is found when the sample is observed through the atomic filter. This ensures that only photons near or on the sodium resonance are detected. A lateral scan under 0–1 excitation is displayed in Fig. 3b. This is a convenient way to limit the detection window to photons on the atomic line.

Earlier studies have proved that the emitted light under excitation into a higher vibronic band is consistent with the linewidth of the 0–0 transition under resonant excitation[10]. The molecular photons therefore resemble the measured linewidths of 12.5–25 MHz in fluorescence excitation. This represents an excellent match for the 9.8-MHz natural linewidth of atomic sodium. In comparison to other studies using quantum dots, which aimed for an optical interaction between single solid-state emitters and atomic vapour[20], this linewidth represents a reduction by more than three orders of magnitude. In a Hanbury–Brown and Twiss experiment (HBT), we analyse the single-photon nature of the emitted photons. Antibunching, measured as the second-order correlation function, $g^{(2)}(\tau)$, shows values down to 0.07 for zero time delay, with no further Rabi oscillations, which are always present under high-power 0–0 excitation. Figure 4a shows that up to 6×10^5 c.p.s. can be detected from a single molecule using a solid immersion lens. This flux represents an increase compared to earlier studies[20] by orders of magnitude.

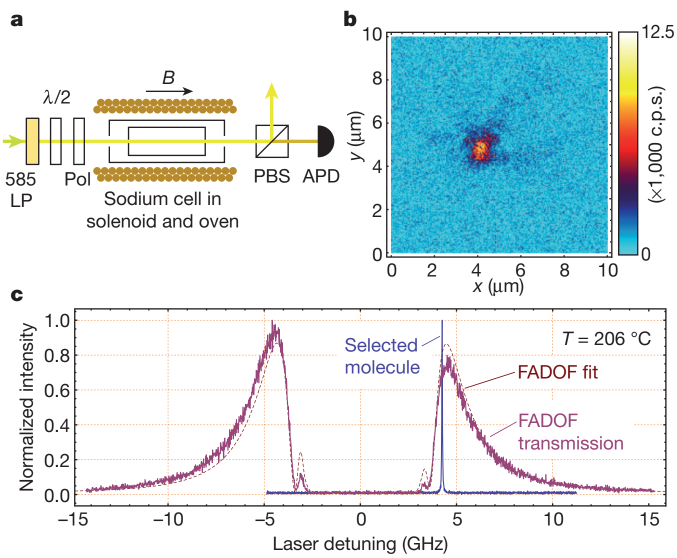


Figure 3 | Narrow-band filtering of the molecular emission by atomic vapour. **a**, Experimental configuration for the sodium Faraday filter. (Pol, polarizer; PBS, polarizing beam splitter; APD, avalanche photo diode). **b**, Lateral scan across the sample, with detection based solely on the near-resonant photons. **c**, Transmission spectrum of the Faraday filter and the selected molecule for this study (FADOF, Faraday anomalous dispersion optical filter).

To demonstrate the usability of the photons generated, we introduce an experiment that simultaneously reveals the single-photon nature of the source and its narrowband spectral match to the sodium line. After selecting a molecule in the vicinity of the Doppler broadening of the atomic transition (Fig. 4b), we observe the emission in an HBT experiment (Fig. 4c). The dispersion of the atomic vapour close to the optical transition is large, so the group velocity v_g of light passing the medium is significantly reduced. The single photons are slowed down in the atomic medium. We unbalanced the HBT configuration by placing a cell containing hot atomic sodium vapour into one arm; the zero point of the coincidence is shifted and slow light on the sodium D_2 line is observed. The measured group velocity delay amounts to 6 ns. A straightforward calculation of the refractive index $n = n' + in''$ (where i is $\sqrt{-1}$) and of the resulting group velocity v_g, using the formula:

$$v_g = \frac{c}{n + \omega \frac{dn}{d\omega}} \tag{1}$$

(with c as the speed of light in vacuum, n as the refractive index, and ω as the angular frequency of the photons) gives a perfect match of 6 ns for the 210 °C hot vapour cell (Fig. 4b, purple line). The transmission intensity through the cell is reduced by a factor of two. The slow light is observable only with narrowband single photons that match the sodium D line. No significant change, or asymmetry in the actual shape of the $g^{(2)}(\tau)$ function, is observed when the line is shifted. This experiment can only be performed with narrowband single photons, because otherwise the resulting $g^{(2)}(\tau)$ function would display different arrival times for differently detuned photons, and would therefore broaden the antibunching signal, while reducing the dip height. In Fig. 4e the calculated $g^{(2)}(0)$ visibility versus the spectral width of the photons is depicted. This emphasizes that the linewidth of the molecular emission and the atomic line match, and that the linewidth reduction by three orders of magnitude in comparison to earlier experiments is a relevant step forward[20].

In these experiments, we combine the unsurpassed brightness and the narrowband nature of light emitted by a single organic molecule with the spectral properties of an atomic system. In cryogenic experiments, the reported use of an atomic vapour filter in single-emitter detection leads to an enhancement of the signal-to-noise ratio of more than 20%. It can be extended to the broad emission band in room-temperature studies and allows for more detectable photons in single-emitter spectroscopy. Photons that are on and close to an atomic resonance allow for an optical interaction with a stationary qubit of a different kind. The high interaction efficiencies in single-pass configurations[6,21] mean that the photons originating from atoms can efficiently interact with single

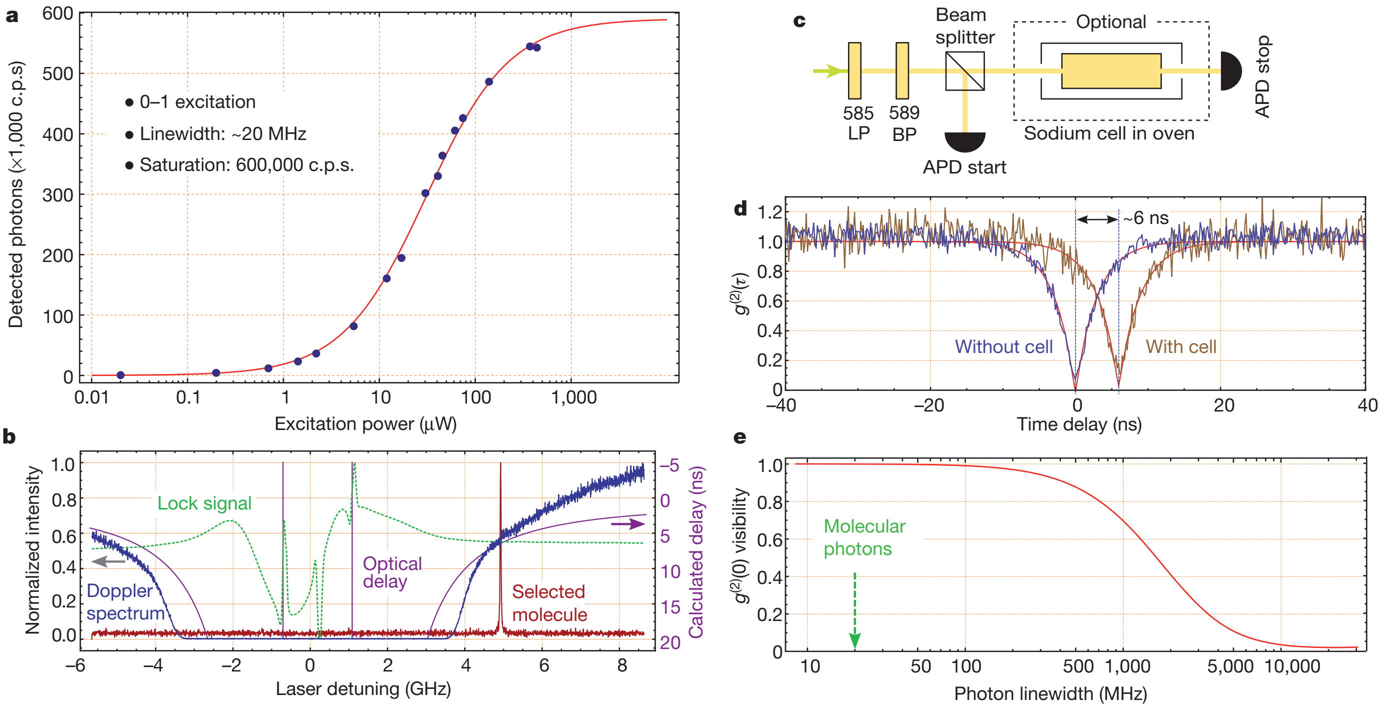


Figure 4 | Near-sodium-resonance photons. **a**, Saturation curve for the molecular single photons, using a solid immersion lens. **b**, By resonant excitation (0–0 excitation) the zero-phonon emission of the molecule under study can be identified. The molecule (red line) was chosen such that a fair amount of photons is transferred through the sodium cell (see the Doppler broadened absorption spectrum; blue line). For a specific detuning, the group velocity delay of single photons is depicted (purple line) **c**, By excitation of the single molecule into a higher vibrational state, the generated photons are analysed in a HBT-type experiment. Optionally (dashed box), a hot atomic vapour cell can be placed into one arm of the setup. To achieve further rejection of the excitation laser, a 585-nm long-pass (LP) filter and a 589-nm band-pass (BP) filter were introduced. **d**, The antibunching was recorded with no cell present in the optical path (blue line). When the hot sodium cell is inserted into the optical setup, the photons are passing through the cell with a reduced effective group velocity v_g. Given that the optical path is effectively longer due to the increased refractive index, the zero position of intensity autocorrelation function is shifted by 6 ns (brown line). **e**, Calculated visibility of the recorded auto-correlation function, detected through an atomic vapour cell, versus single-photon linewidth.

molecules and vice versa. The narrowband nature of the light emitted means that the single photons presented can be optimally linked to ultracold atomic gases. So far, realizations of storage and entanglement protocols have relied on probabilistic photon sources. The high-flux and narrow-bandwidth single-photon sources presented here will boost the available efficiencies in such schemes. Experiments in the fields of Rydberg-mediated single-photon nonlinearities[7,8,22,23] suggest that the realization of multiple-photon quantum gates is in reach. To produce Fock states with higher photon numbers, several molecules can be electrically tuned to a desired transition, and triggered by optical π-pulses[24]. Given that the emission stems from a volume that is in all dimensions much smaller than the optical wavelength, and collection efficiency can approach unity[3], the resulting multi-photon state will share one optical mode.

The experiments in an unbalanced HBT configuration prove that the photons generated can be stored in an atomic quantum memory. For quantum communication in daylight[25], the generated sodium-resonant photons can be transmitted with minimal solar disturbance, because the sodium-D-line transitions exhibits one of the strongest Fraunhofer lines in sunlight. The DBT molecule allows for an optical interface with potassium and rubidium. Other molecules that exhibit an overlap to a variety of atomic systems are also presented. We believe this work paves the way to the discovery of many simultaneously bright and narrowband single-photon sources, and that a picowatt of continuous single-photon emission matched to an atomic system will soon be exceeded.

METHODS SUMMARY

The experiments were performed in a home-build low-temperature confocal microscope with a detection channel arranged in a HBT configuration. One path of the HBT interferometer was 30 cm longer, where a conventional interference filter or the hot atomic cell can be placed. The excitation source was a single-mode rhodamine 6G dye ring laser. Its frequency was stabilized by Doppler free resonances of atomic vapour. This atomic vapour cell also served as a frequency reference, when the laser frequency was swept to record the fluorescence spectra of the molecules. Single-molecule samples for spectroscopy with sodium vapour were made by preparing a saturated solution of DBATT in n-tetradecane followed by appropriate dilution. The samples of molecules for experiments with rubidium and potassium were produced by melting the naphthalene with admixture of DBT molecules between two cover slips on a hot plate. For experiments, all samples were shock-frozen. The microelectrodes for Stark shift control were produced by the lithography method on the glass cover slide that served as the lower single-molecule sample substrate. The upper part of the sample was either another cover slip or a cubic zirconia solid immersion lens with a diameter of 3 mm. The atomic vapour cells were home-built and placed inside a heater combined with a solenoid. For the filtering and slow-light experiments the number of atoms per volume, and hence optical density, was controlled by temperature.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 November 2013; accepted 27 February 2014.

1. Kimble, H. J. The quantum internet. *Nature* **453,** 1023–1030 (2008).
2. Nisbet-Jones, P. B. R., Dilley, J., Ljunggren, D. & Kuhn, A. Highly efficient source for indistinguishable single photons of controlled shape. *New J. Phys.* **13,** 103036 (2011).
3. Lee, K. *et al.* A planar dielectric antenna for directional single-photon emission and near-unity collection efficiency. *Nature Photon.* **5,** 166–169 (2011).
4. Lounis, B. & Orrit, M. Single-photon sources. *Rep. Prog. Phys.* **68,** 1129–1179 (2005).
5. Hettich, C. *et al.* Nanometer resolution and coherent optical dipole coupling of two individual molecules. *Science* **298,** 385–389 (2002).
6. Wrigge, G., Gerhardt, I., Hwang, J., Zumofen, G. & Sandoghdar, V. Efficient coupling of photons to a single molecule and the observation of its resonance fluorescence. *Nature Phys.* **4,** 60–66 (2008).

7. Dudin, Y. O. & Kuzmich, A. Strongly interacting Rydberg excitations of a cold atomic gas. *Science* **336,** 887–889 (2012).
8. Firstenberg, O. *et al.* Attractive photons in a quantum nonlinear medium. *Nature* **502,** 71–75 (2013).
9. Boiron, A.-M., Lounis, B. & Orrit, M. Single molecules of dibenzanthanthrene in *n*-hexadecane. *J. Chem. Phys.* **105,** 3969–3974 (1996).
10. Lettow, R. *et al.* Realization of two Fourier-limited solid-state single-photon sources. *Opt. Express* **15,** 15842–15847 (2007).
11. Petelski, T., Fattori, M., Lamporesi, G., Stuhler, J. & Tino, G. Doppler-free spectroscopy using magnetically induced dichroism of atomic vapor: a new scheme for laser frequency locking. *Eur. Phys. J. D* **22,** 279–283 (2003).
12. Wild, U. P., Güttler, F., Pirotta, M. & Renn, A. Single molecule spectroscopy: Stark effect of pentacene in *p*-terphenyl. *Chem. Phys. Lett.* **193,** 451–455 (1992).
13. Orrit, M., Bernard, J., Zumbusch, A. & Personov, R. Stark effect on single molecules in a polymer matrix. *Chem. Phys. Lett.* **196,** 595–600 (1992).
14. Nicolet, A. A. L., Hofmann, C., Kol'chenko, M. A., Kozankiewicz, B. & Orrit, M. Single dibenzoterrylene molecules in an anthracene crystal: spectroscopy and photophysics. *ChemPhysChem* **8,** 1215–1220 (2007).
15. Pirotta, M., Renn, A., Werts, M. H. & Wild, U. P. Single molecule spectroscopy. Perylene in the Shpol'skiĭ matrix *n*-nonane. *Chem. Phys. Lett.* **250,** 576–582 (1996).
16. Nonn, T. & Plakhotnik, T. Fluorescence excitation spectroscopy of vibronic transitions in single molecules. *Chem. Phys. Lett.* **336,** 97–104 (2001).
17. Kiraz, A., Ehrl, M., Bräuchle, C. & Zumbusch, A. Low temperature single molecule spectroscopy using vibronic excitation and dispersed fluorescence detection. *J. Chem. Phys.* **118,** 10821–10824 (2003).
18. Dick, D. J. & Shay, T. M. Ultrahigh-noise rejection optical filter. *Opt. Lett.* **16,** 867–869 (1991).
19. Harrell, S. D. *et al.* Sodium and potassium vapor Faraday filters revisited: theory and applications. *J. Opt. Soc. Am. B* **26,** 659–670 (2009).
20. Akopian, N., Wang, L., Rastelli, A., Schmidt, O. G. & Zwiller, V. Hybrid semiconductor-atomic interface: slowing down single photons from a quantum dot. *Nature Photon.* **5,** 230–233 (2011).
21. Tey, M. K. *et al.* Strong interaction between light and a single trapped atom without the need for a cavity. *Nature Phys.* **4,** 924–927 (2008).
22. Pritchard, J. D. *et al.* Cooperative atom-light interaction in a blockaded Rydberg ensemble. *Phys. Rev. Lett.* **105,** 193603 (2010).
23. Parigi, V. *et al.* Observation and measurement of interaction-induced dispersive optical nonlinearities in an ensemble of cold Rydberg atoms. *Phys. Rev. Lett.* **109,** 233602 (2012).
24. Gerhardt, I. *et al.* Coherent state preparation and observation of Rabi oscillations in a single molecule. *Phys. Rev. A* **79,** 011402 (2009).
25. Peloso, M. P., Gerhardt, I., Ho, C., Lamas-Linares, A. & Kurtsiefer, C. Daylight operation of a free space, entanglement-based quantum key distribution system. *New J. Phys.* **11,** 045007 (2009).
26. Mais, S., Basche, T., Mueller, G., Müllen, K. & Bräuchle, C. Probing the spectral dynamics of single terrylenediimide molecules in low-temperature solids. *Chem. Phys.* **247,** 41–52 (1999).
27. Kiraz, A., Ehrl, M., Bräuchle, C. & Zumbusch, A. Ultralong coherence times in the purely electronic zero-phonon line emission of single molecules. *Appl. Phys. Lett.* **85,** 920–922 (2004).
28. Jelezko, F., Tamarat, P., Lounis, B. & Orrit, M. Dibenzoterrylene in naphthalene: a new crystalline system for single molecule spectroscopy in the near infrared. *J. Phys. Chem.* **100,** 13892–13894 (1996).
29. Kummer, S., Basche, T. & Bräuchle, C. Terrylene in *p*-terphenyl: a novel single crystalline system for single molecule spectroscopy at low temperatures. *Chem. Phys. Lett.* **229,** 309–316 (1994).
30. Kummer, S. *et al.* Absorption, excitation, and emission spectroscopy of terrylene in *p*-terphenyl: bulk measurements and single molecule studies. *J. Chem. Phys.* **107,** 7673–7684 (1997).

Acknowledgements We thank W. Kiefer for the calculation of the sodium-D_2-FADOF transmission (Fig. 3). G.S. acknowledges support by J. Pflaum (University of Würzburg). I.G. acknowledges discussions with R. Löw and S. Hofferberth (University of Stuttgart). J.W. acknowledges support by the Max Planck Society (via a Max Planck fellowship), the BMBF (via the projects QuORep and Q.com) and the EU (via the project SIQS and the ERC grant SQUTEC).

Author Contributions I.G. conceived the idea. P.S., G.S. and I.G. prepared and conducted the experiments. I.G. and J.W. supervised the team and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.G. (i.gerhardt@fkf.mpg.de).

METHODS

Sample preparation. To prepare the single-molecule samples for the cryogenic experiments in conjunction with sodium, a saturated solution of DBATT (PAH Research, Greifenberg) in n-tetradecane (Aldrich) was produced. For the samples below 0–0 excitation this solution was further diluted by a factor of 50. The sample was sandwiched between two cover slides and shock-frozen in a helium-bath cryostat (Janis). The imaging optics consists of a 63× objective (Melles Griot, numerical aperture 0.85), which withstands the cryogenic environment. Single molecules were identified by confocal microscopy. The incident laser power was usually 10 nW, measured before the laser enters the cryostat. We emphasize that we carefully ensured the same conditions when changing the filters from the commercial filter and the sodium vapour filter. The only step was a careful optimization of the avalanche photo diode, to compensate for eventual beam shifts caused by the vapour cell windows.

The samples for experiments with potassium and rubidium were prepared by melting DBT (PAH Research, Greifenberg) and naphthalene together, resulting in a light-green molten bath. For sample preparation, some crumbs of this mixture were melted on a hot plate between two clean cover slides and then shock-frozen on a metal plate. The sample would not be suitable for single-molecule studies in the centre of the inhomogeneous broadening, because the concentration would be too high. In general, molecules are very sparse far from the inhomogeneous broadening, reported to be centred around 757.7 nm (ref. 28). Therefore, Fig. 2e shows the spectra of multiple single molecules at different locations in the sample, recorded separately. The cryogenic experiment was the same as with the sodium experiments, except that the commercial long-pass filter was changed to a 780-nm transition edge (IDEX/Semrock), and the laser was changed to a near-infrared configuration.

For the 0–1 transition experiments, the solution for the 0–0 experiments was further diluted by a factor of 1,000. This is necessary because the transition linewidth is approximately 50 GHz, and thus more than 1,000 times broader than the 0–0 transition. The excitation into a higher vibronic band requires a higher transition energy and the laser had to be blue-detuned by approximately 245.2 cm^{-1}. To acquire the saturation curve, the experimental configuration was not changed except that the sample solution was sandwiched between a cover slide and a hemispherical solid immersion lens with radius 1.5 mm made of cubic zirconia ($n = 2.18$, A.W.I. Industries).

The recorded auto correlation functions were acquired in a start-stop type fashion, using a specialized card for time-correlated single-photon counting (SPC730, Becker and Hickl).

The atomic vapour cell was made of quartz glass, produced in-house, and had an optical length of 100 mm for the filtering experiments and 72 mm for the studies on slow light. Both cells contained more sodium than necessary for a saturated vapour. For the filtering experiments, the temperature of the cell was 155 °C and the optical density was 5. In principle, the optical rejection can be significantly higher, but it is limited by finite scattering of resonant photons within the cell and towards our single-photon detector, placed about 200 mm behind the cell. The measured optical suppression at 1 mW for a vapour cell at 200 °C is higher than an absorbance of 6 at the centre of the Doppler-broadened line (589.1583 nm in vacuum).

For the slow-light experiments the cell was used in a triple pass configuration to increase the optical path length, thus increasing the delay of the light to enhance measurability. For the FADOF type experiments, the magnetic field was limited to approximately 10 mT. To achieve high transmission, the temperature was adapted. In principle, it is possible, by increasing the magnetic field by a factor of 18, to achieve a very high transmission very close to the centre of the sodium transition[19]. To block out the excitation light, the 585LP filter for the 0–1 excitation was used, because the excitation power was of the order of milliwatts and the rejection of the two crossed polarizers only of the order of an optical density of five, given that the sodium-cell windows introduce some aberrations to the wavefront.

Data processing. All displayed data are presented as recorded (raw data), with three exceptions: the baseline of the spectra of DBT in naphthalene was shifted to zero, such that the different background level of the molecules (maximum 1,000 c.p.s.) is levelled out. The saturation curve for a single DBATT molecule was corrected for a linear increasing background contribution, at most 10% of the counts. The autocorrelation measurements in Fig. 4 would be influenced by this background contribution with a shifted baseline. The raw data had a baseline of 0.2 using the cell and 0.13 without the cell. Subsequently, the antibunching data was corrected for background, in the following manner:

$$g^{(2)}(\tau) = \frac{G(\tau)_{\mathrm{norm}} + \eta^2 - 1}{\eta^2} \qquad (2)$$

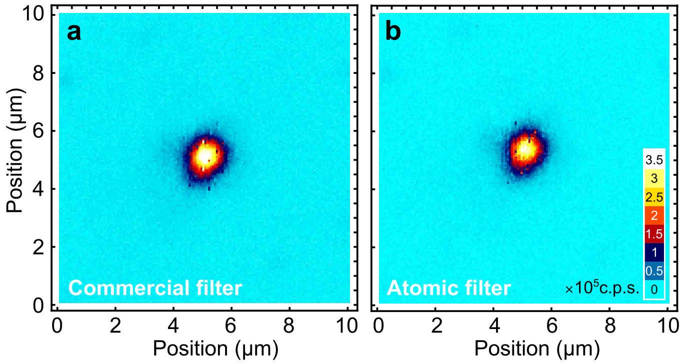
where $G(\tau)_{\mathrm{norm}}$ is the coincidence rate normalized to the Poissonian source, and defined as $\eta = s/(s+b)$, where s is the signal and b the background, as in the literature[31]. The background was estimated by recording a spot 3 μm beside the molecule.

Calculations. To calculate the group velocity delay in atomic vapour, we use the basic mathematics of ref. 32. We do not need to average over the different velocity classes in the hot vapour, because the group velocity is dominated by the Lorentzian contribution, given that the Gaussian part decays exponentially and has no further influence[33]. The estimated group velocity v_g was calculated using equation (1). The refractive index of the atomic medium was calculated as:

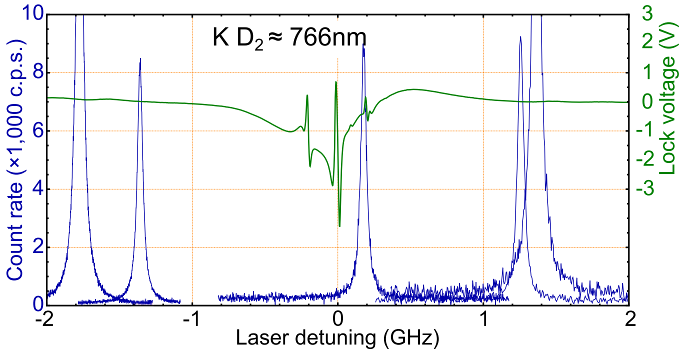
$$n = 1 + \sum_i \frac{Ne^2 f_i}{4\varepsilon_0 m\omega_{0i}} \frac{(\omega_{0i} - \omega)}{(\omega_{0i} - \omega)^2 + \left(\frac{\gamma}{2}\right)^2} \qquad (3)$$

Here i represents the specific transition, f_i the corresponding oscillator strength, ω_0 the centre frequency of the transition, N the number of atoms in the vapour, e and m the electron charge and mass, respectively. γ is the natural linewidth of the atomic transition (≈2π × 9.8 MHz).

31. Kitson, S. C., Jonsson, P., Rarity, J. G. & Tapster, P. R. Intensity fluctuation spectroscopy of small numbers of dye molecules in a microcavity. *Phys. Rev. A* **58,** 620–627 (1998).
32. Boyd, R. W. & Gauthier, D. J. in *Progress in Optics* Vol. 43, Ch. 6, 497–530 (Elsevier, 2002).
33. Shi, Z., Boyd, R. W., Camacho, R. M., Vudyasetu, P. K. & Howell, J. C. Slow-light Fourier transform interferometer. *Phys. Rev. Lett.* **99,** 240801 (2007).



Extended Data Figure 1 | Imaging fluorescent samples, supported by atomic vapour. Single fluorescent bead, imaged in a confocal microscope, illuminated with laser light, locked to the crossover resonance of the sodium D_2 transition. Although the overall count rate does not substantially differ, the signal-to-noise ratio is measured to be 210 when using the commercial filter, and 240 when only the sodium filter is used. Integration time per pixel is 2 ms.



Extended Data Figure 2 | Combined spectroscopy of DBT and potassium. Spectra of single DBT molecules and atomic K vapour on the K D_2 line around 766 nm. The molecules in the sample are sparse. Therefore, the image represents multiple recordings at several lateral positions.

doi:10.1038/nature13189

Spontaneous transfer of chirality in an atropisomerically enriched two-axis system

Kimberly T. Barrett[1], Anthony J. Metrano[1], Paul R. Rablen[2] & Scott J. Miller[1]

One of the most well-recognized stereogenic elements in a chiral molecule is an *sp*[3]-hybridized carbon atom that is connected to four different substituents. Axes of chirality can also exist about bonds with hindered barriers of rotation; molecules containing such axes are known as atropisomers[1]. Understanding the dynamics of these systems can be useful, for example, in the design of single-atropisomer drugs[2] or molecular switches and motors[3]. For molecules that exhibit a single axis of chirality, rotation about that axis leads to racemization as the system reaches equilibrium. Here we report a two-axis system for which an enantioselective reaction produces four stereoisomers (two enantiomeric pairs): following a catalytic asymmetric transformation, we observe a kinetically controlled product distribution that is perturbed from the system's equilibrium position. As the system undergoes isomerization, one of the diastereomeric pairs drifts spontaneously to a higher enantiomeric ratio. In a compensatory manner, the enantiomeric ratio of the other diastereomeric pair decreases. These observations are made for a class of unsymmetrical amides that exhibits two asymmetric axes—one axis is defined through a benzamide substructure, and the other axis is associated with differentially *N*,*N*-disubstituted amides. The stereodynamics of these substrates provides an opportunity to observe a curious interplay of kinetics and thermodynamics intrinsic to a system of stereoisomers that is constrained to a situation of partial equilibrium.

The generation of enantiopure, chiral molecules remains relevant to many scientific fields, from the study of biological systems to materials science. One critical challenge is that enantioenriched compounds are not fully equilibrated ensembles. Enantiopure compounds represent an ensemble of higher free energy owing to the entropic penalty associated with a one-state, homochiral composition relative to the corresponding two-state racemate[4]. The thermodynamic benefit of a two-state system can counteract asymmetric synthetic efforts, which are often performed under kinetic control, in pursuit of single-enantiomer compounds. Examples of racemization include the isomerization of an enantioenriched α-chiral aldehyde (Fig. 1a) and thermal equilibration of an axially chiral biaryl compound (Fig. 1b)[5].

Atropisomerization, the phenomenon of equilibration of stereoisomers about a rotational axis, is an issue of contemporary interest in organic[6], materials[7,8] and medicinal chemistry[9]. As part of a programme targeted at developing catalysts that produce unique atropisomers selectively, we recently discovered that catalyst **1** is effective for the selective synthesis of enantiomerically enriched benzamides, converting racemic compounds like **2** into the corresponding tribromides (**3**), with enantiomer ratios of up to 94:6 (yields up to 89%; Fig. 2)[10]. By virtue of a low barrier to racemization of the substrate (**2**), a peptide-catalysed dynamic kinetic resolution[11] allows for the preferential bromination of one enantiomer, leading to restricted bond rotation and stable atropisomers (**3**) at room temperature for extended periods.

A situation of greater stereochemical intrigue is established when one considers a compound such as **4**. Benzamide **4** (Fig. 3), with two different substituents on the amide *N*-atom, may exist as four different stereoisomers (enantiomeric pairs of both ***cis*-4** and ***trans*-4**). Thus, we wondered if it might be possible to identify catalysts that select not only for individual enantiomers, but also for individual diastereomers—each of the four possible stereoisomers (**5**). Whereas peptides such as **1** were envisioned to provide stereochemical control over the atropisomeric axis of **4**, it was unclear at the onset of this study whether control of the amide axis disposition (*cis* amide versus *trans* amide) could be accomplished with the same catalyst. If interconversion among all possible diastereomers of the two-axis starting material **4** were possible (with low barriers to isomerization within the starting materials)[12], one could envision four unique catalysts that might accomplish the task. Of course, a critical issue is the overall stability of the individual stereoisomeric products (variants of **5**). Low barriers to rotation about either the benzamide axis (Ar–CO, where Ar is aryl; red bond), the amide bond axis (C–N, blue bond)[13], or both, in a concerted manner[14,15], could conspire to erode kinetic selectivity.

Our studies provided an opportunity to observe a curious result. When ***rac*-4** is exposed to dibromodimethylhydantoin (DBDMH) in the absence of a chiral catalyst, under conditions otherwise analogous to those of Fig. 2 (−40 °C), the expected racemic products are formed over the course of about 50 h (70% yield), as a mixture of four stereoisomers (**5**). After the reaction is quenched, the phenol is converted to the methyl ether for analytical purposes to generate **5-(Me)**, where Me is methyl. When the isomeric mixture is purified and analysed by chiral high-performance liquid chromatography (HPLC) (about 1 h after quenching, at about 25 °C), the first measurement reveals a ratio of 40:60 ***trans*-5-(Me)**:***cis*-5-(Me)** isomers, each in racemic form (Fig. 4a). If the sample is allowed to stand at room temperature (dissolved in 10% *i*PrOH/hexanes, where *i*Pr is *iso*-propyl) and is reanalysed at a much later time point (50 h), the

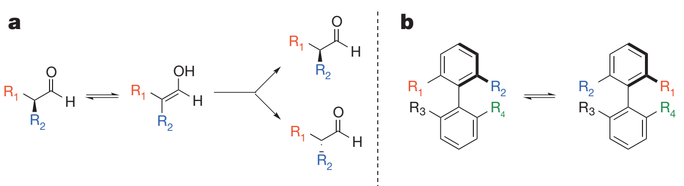


Figure 1 | Stereochemical interconversion of chiral organic compounds. **a**, Racemization of an enantioenriched α-substituted aldehyde. **b**, Atropisomerization of an axially chiral biaryl compound.

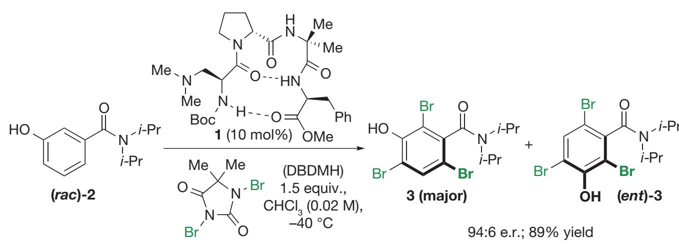


Figure 2 | Catalytic enantioselective bromination of *N*,*N*-diisopropyl benzamides. e.r., enantiomeric ratio; *i*-Pr, *iso*-propyl; Boc, *tert*-butoxycarbonyl; *rac*, racemic; *ent*, enantiomer; $CHCl_3$, chloroform.

[1]Department of Chemistry, Yale University, PO Box 208107, New Haven, Connecticut 06520-8107, USA. [2]Department of Chemistry and Biochemistry, Swarthmore College, Swarthmore, Pennsylvania 19081-1397, USA.

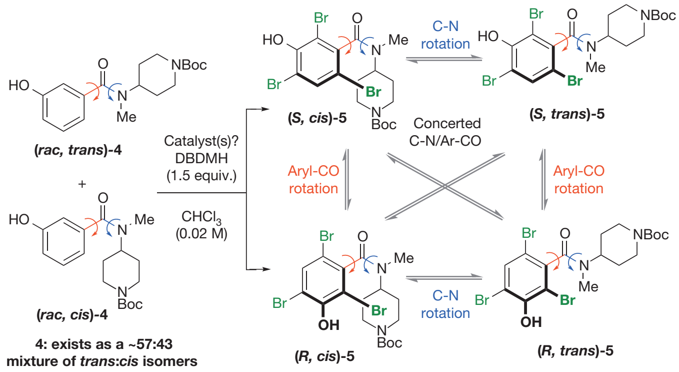


Figure 3 | Proposed catalytic enantioselective bromination of a two-axis, differentially substituted benzamide. The use of the *R*- and *S*-stereochemical descriptors are in accord with convention, and are defined interchangeably with the also-used *M*- and *P*-stereochemical convention[16]: *R* = *M*; *S* = *P*.

trans:*cis* ratio is observed to increase to 76:24 (Fig. 4b). Notably, although the *cis*-amide of **4** is the minor component of the starting material over a wide temperature range, including at the reaction temperature of −40 °C, the *cis*-amide of **5** (assayed as **5-(Me)**) appears to be generated in slight excess. Thus, it is apparent that there is some modest kinetic selectivity for the *cis*-isomer, which equilibrates at room temperature to the thermodynamically more stable *trans*-amide over time. It is notable that the amide of **5-(Me)** exhibits a barrier to C–N bond isomerization that is high relative to typical amides, but still too low to be effectively arrested at room temperature. The experimental and calculated barriers to C–N bond rotation are determined and discussed below.

When the reaction is performed in the same manner with chiral catalyst **1** (94% yield, within 24 h), a more elaborate scenario is observed. When the product mixture is analysed by chiral HPLC at the first time point (about 2 h after quenching, at 25 °C), a *trans:cis* ratio of 43:57 is observed (Fig. 4c). In this measurement, the *trans*-amide enantiomeric ratio is 66:34, whereas the *cis*-amide enantiomeric ratio is recorded as 88:12. As the sample is allowed to stand in solution (10% *i*PrOH/hexanes)—long after the chiral catalyst has been removed from the system—the following changes occur spontaneously in the product distribution. At 10 h after quenching (Fig. 4d), the *trans*:*cis* ratio moves to 54:46, enhancing the population of the *trans*-isomer as expected. In parallel, the enantiomeric ratios of both amide isomers change as the system moves towards the *trans:cis* amide equilibrium position, with the *cis*-amide enantiomeric ratio decreasing to 86:14, while the *trans*-amide enantiomeric ratio spontaneously increases to 72:28. These changes continue as the system continues to equilibrate. At 24 h (Fig. 4e), the

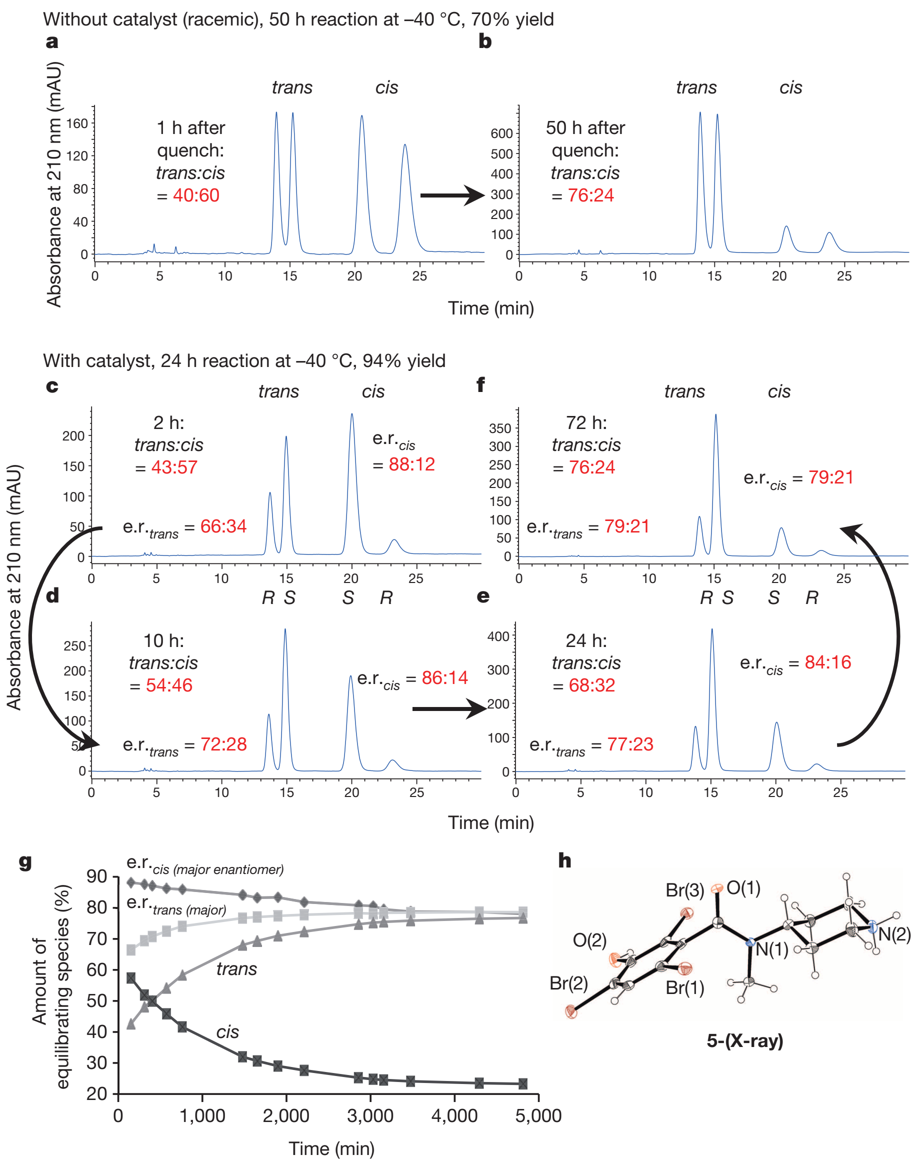


Figure 4 | Experimental data describing the stereochemical behaviour of the isomeric benzamide products. a–f, Chiral HPLC traces of **5-(Me)** analysed at room temperature. **a, b,** Reactions run at −40 °C in the absence of catalyst. Time zero is defined as the point of reaction quench, followed by purification and HLPC analysis at 25 °C. **c–f,** Reactions run in the presence of catalyst and subsequently monitored over time after reaction work-up. Peak assignments in order of elution: peak 1: *R, trans*; peak 2: *S, trans*, peak 3: *S, cis*; peak 4: *R, cis*. **g,** Graphical representation of changes in isomeric components. **h,** Crystallographic structure of (***S, trans***)-derivative used for the absolute stereochemistry assignments[16]. mAU, milliabsorbance units.

trans:*cis* ratio is 68:32; the *cis*-amide enantiomeric ratio erodes to 84:16, and the *trans*-amide enantiomeric ratio increases to 77:23. After 72 h, the product ratios have stabilized (Fig. 4f), and the apparent equilibrium position of the amide diastereomers has been reached, with a *trans*:*cis* ratio of 76:24. At this stage, the *cis*-amide enantiomeric ratio is 79:21, identical to the *trans*-amide enantiomeric ratio. These observations are depicted graphically in Fig. 4g. Although data are shown in Fig. 4 for reactions conducted at $-40\,^{\circ}$C, the observations are qualitatively reproduced when the experiments are repeated at several different temperatures (see Supplementary Information section VIIa).

Our observations reflect a situation of spontaneous enantioenrichment for one of the product diastereomers (*trans*), with a compensatory decrease in enantiomeric ratio for the other diastereomer (*cis*). Interestingly, a 50:50 racemic mixture is not observed, with the system retaining enantioenrichment even after prolonged periods of time at room temperature, a consequence of the two-axis system failing to reach complete equilibrium within the time frame analysed (see below).

Both the *trans* and the *cis* isomers of the products could be separated by silica gel chromatography—an unusual circumstance—and each produced the expected diastereomeric *trans*:*cis* ratio (76:24) upon standing in solution (10% *i*PrOH/hexanes). In these cases, the enantiomeric ratio of the samples remains virtually constant (*cis*, 87:13; *trans*, 74:26; see Supplementary Information sections VIc and VId and Supplementary Fig. 7 for details).

Through separation of the enantioenriched *cis*- and *trans*-amide isomers, enantioenriched *trans*-material was obtained and provided crystallographic quality material to assign **(*S*, *trans*)-5** as the major isomeric component after amide equilibration (Fig. 4h; see also Supplementary Information section IV). We note that the stereoisomer obtained for assignment of the *S*-*trans* configuration was derived from isolation of the *cis* product, reflecting the crystallization of the major *trans* diastereomer, as *cis*-to-*trans* equilibration occurs over the course of the crystallization experiment. Parenthetically, the absolute configuration of this *S*-*trans* sample, derived from the isolation of enriched *cis* sample, is the same absolute configuration observed when catalyst **1** operates on substrate **2**, to deliver enantioenriched (***S*-3**) with a 94:6 enantiomeric ratio (Fig. 2)[10]. Accordingly, these data provide circumstantial support for the equilibration of the *S*-*cis* isomer to the *S*-*trans* isomer (and also of *R*-*cis* to *R*-*trans*) without interconversion of the axis of chirality. Further details of this scenario are now considered.

As *cis*-to-*trans* amide isomerization proceeds and the diastereomeric ratio reaches its equilibrium position, the final enantiomeric ratios for both the *trans*-amide isomers and the *cis*-amide isomers emerge as equivalent. As is implicit in Fig. 4g, the sum of the major enantiomers of the amide diastereomers ((***S*-*cis***)-**5-(Me)** + (***S*-*trans***)-**5-(Me)**), divided by the sum of the minor enantiomers of the diastereomeric amides ((***R*-*cis***)-**5-(Me)** + (***R*-*trans***)-**5-(Me)**) is near 79:21 (3.76 ± 0.41) at each time point in Fig. 4 (Supplementary Table 6 and Supplementary Fig. 5 contain expanded data sets detailing this point). The convergence of the enantiomeric ratio for both the *trans*- and *cis*-amide diastereomers is consistent with a mechanistic model wherein amide isomerization occurs through independent C–N bond rotation at ambient temperature (Fig. 3), while the enantiomer-defining axis of chirality (Ar–CO) is essentially fixed. The equilibration of the amide isomers, without interconversion of the Ar–CO bond axis, leads to fluctuation of the starting *cis*-amide enantiomeric ratio downward, and the *trans*-amide enantiomeric ratio upward, until amide isomerization achieves the equilibrium ratio, and the enantiomeric ratios of the diastereomers are equivalent.

However, when the *cis*- or *trans*-amide diastereomers are separated by chromatography, and the isolated amide diastereomers are allowed to re-equilibrate, the enantiomeric ratio remains constant in each series. This situation is once again a manifestation of Σ(*S*, *cis* + *S*, *trans*)/Σ(*R*, *cis* + *R*, *trans*) remaining constant. In this case, there is no reservoir of the other amide diastereomer, of a different enantiomeric ratio, to distribute its population of either *S*- or *R*-configuration differentially to the two amide diastereomers at the C–N bond equilibrium position.

To explore the plausibility of these assertions, we complemented our experiments with a series of density functional theory calculations to ascertain the barriers associated with the critical modes of isomerization. The dynamics described above for amide isomerization correspond to experimentally derived free-energy barriers[17,18] of 24.8 kcal mol^{-1} (*trans*-to-*cis*) and 24.1 kcal mol^{-1} (*cis*-to-*trans*) (Fig. 5; see Supplementary Information section VIII). Computations[19,20] provided free-energy barriers of 24.4 kcal mol^{-1} (*trans*-to-*cis*) and 24.0 kcal mol^{-1} (*cis*-to-*trans*) (Supplementary Information section XIII), on a par with both the experimentally determined values and literature values for somewhat related compounds[13]. Additionally, during the entire computed C–N rotation process, the Ar–CO dihedral angle remains close to its value in the ground state, even as the amide axis rotates out of conjugation with the carbonyl and pyramidalizes (**TS-5a**, Fig. 5b). Thus, independent C–N bond rotation appears to have a much lower barrier than any putative process involving a coupled rotation of the Ar–CO axis. This is consistent with the observation that erosion of the overall enantioenrichment of the system does not occur at room temperature[21,22].

Of interest to the present system are previous reports by Clayden, which showed that rotations about the Ar–CO bond axis of sterically hindered tertiary amides follow a mechanistic course involving concerted rotations of the Ar–CO axis and the amide C–N bond axis, in a gearing fashion, when sufficient energy is available to the system[21,23]. Our calculations for the present system reassert these conclusions. However, the system we present here is distinct in that (1) the action of a chiral catalyst delivers diastereomeric amides of different enantiomeric ratio that allows for the observation of fluctuating enantiomeric ratios, and

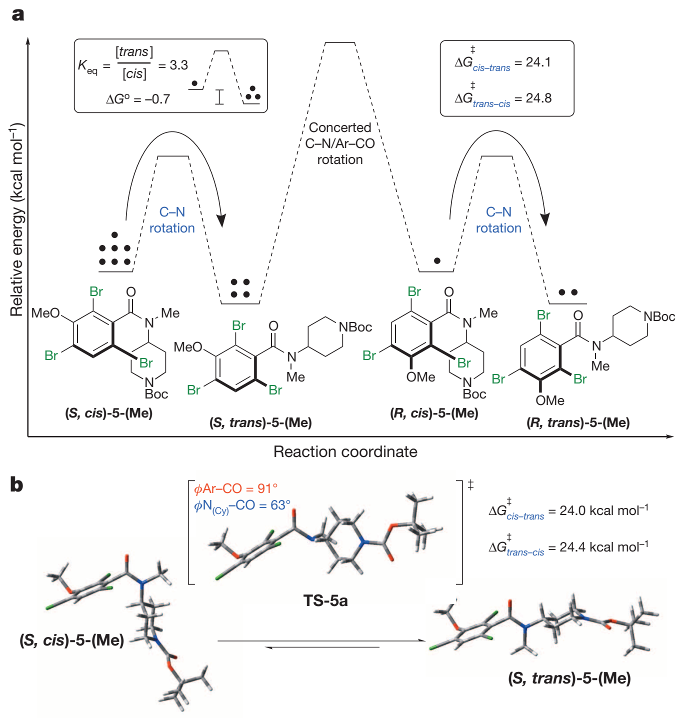


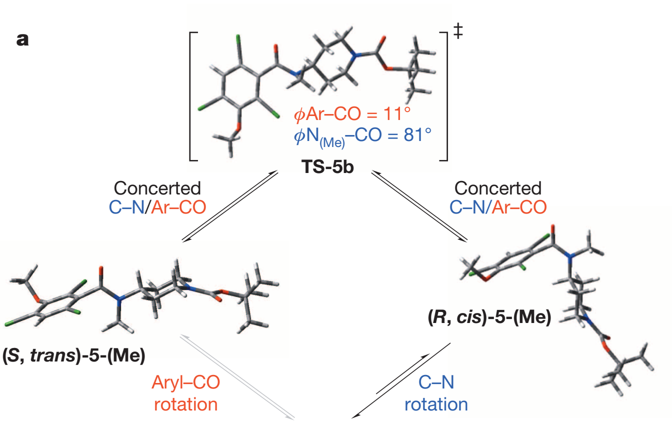
Figure 5 | Energetic considerations and analysis of the stereoisomerizations. a, Energy diagram representing the initial experimental isomeric populations, equilibrium populations (inset 1), and experimentally derived amide rotational barriers (inset 2) of **5-(Me)** at 25 °C (10% *i*PrOH/hexanes). K_{eq} is the equilibrium constant, Δ*G* is the Gibbs free energy and ϕ represents the dihedral angle. **b**, Computed ground states and transition state (TS) geometries for amide isomerization of (***S***)-**5-(Me)**. Computations were performed with a torsional potential energy scan of the C–N dihedral angle, followed by geometry optimization of all stationary points (transition states and minima) using B3LYP/6-31+G(d,p) in the Gaussian 09 software (http://www.gaussian.com). Harmonic vibrational frequencies were calculated at the same level of theory to determine free energies (Δ*G*), and single point energies were computed using M06-2X/6-311++G(2d,3p). Cy, cyclic.

(2) there are distinct steric demands of the substrate that separate the energetic barriers of geared Ar–CO/C–N isomerization from independent C–N bond rotation substantially. Computations employing a relaxed potential energy scan of the Ar–CO dihedral angle led to a simultaneous rotation about the amide C–N bond. The optimized transition states along this torsional energy profile were marked by non-coplanar *N*- and aryl-substituents and imaginary frequencies that showed coupled rotation about both axes. Compared to independent Ar–CO rotation, which suffers from an implausibly high computed barrier, this concerted Ar–CO/C–N rotation represents the lowest-energy pathway to inversion of the atropisomeric axis (**TS-5b**, Fig. 6a). However, below these energetic thresholds, our results with compound **5** are consistent with independent C–N rotation, as noted above.

Although amide isomerization occurred at ambient temperatures, we could induce racemization only through heating the atropisomeric benzamides at higher temperatures (toluene, >60 °C). We determined the free-energy barrier to racemization experimentally[4], and found a value of 27.8 kcal mol^{-1} (at 70 °C) for the ***cis*-5-(Me)** isomer, and a value of 28.6 kcal mol^{-1} for the lower energy ***trans*-5-(Me)** isomer (Fig. 6b; see Supplementary Information section IX). When the barriers were computed, values of 28.2 kcal mol^{-1} and 28.5 kcal mol^{-1} for ***cis*-5-(Me)** and ***trans*-5-(Me)** were found, respectively, in good accord with the experimentally determined values and literature precedent (see Supplementary Information section XIV)[24]. Taken together, our results imply that the racemization pathway involves, for a given amide diastereomer (for example, **(*S, trans*)-5-(Me)** to **(*R, trans*)-5-(Me)**), first, a concerted motion of the Ar–CO/C–N axes to convert **(*S, trans*)-5-(Me)** to **(*R, cis*)-5-(Me)**; and, second, an independent motion of the C–N axis converting **(*R, cis*)-5-(Me)** to **(*R, trans*)-5-(Me)**, as the original amide diastereomer is restored (Fig. 6a). From the experiments and calculations above, it appears that the concerted two-axis rotation operates at increased temperature, and exhibits a high enough barrier that it is prohibitively slow at room temperature.

In summary, we have observed a stereoisomeric system in which spontaneous enantiomeric enrichment occurs in homogeneous solution, with a compensatory erosion of enantiomeric ratio in a coupled diastereomer. The observation is made possible by the formation of a nonequilibrium mixture of amide diastereomers, wherein each enantiomeric pair is produced under the kinetic influence of a chiral catalyst. Amide isomerization occurs with fluctuation of each diastereomer's enantiomeric ratio as *cis*–*trans* amide equilibration occurs, yet the chirality-defining element does not enantiomerize. Instead, the overall enantioenrichment of the system is retained as the populations of each isomer interconvert. Our understanding of the dynamic processes includes an assessment of the intrinsic barriers for the isomerizations, as a function of individual bond rotations, or those that may occur in a concerted manner. The interconversion of stereoisomers in the presence of a fixed element of chirality has been exploited to great advantage in the development of asymmetric reactions[25,26]. Yet, the present case of fluctuating enantiomeric ratios, without epimerization of a chiral element within reaction products that have been isolated away from their equilibrium positions, is distinct. These features may be of interest given the current literature on atropisomerization[27,28]. Moreover, these observations may inform endeavours where spontaneous transfer of chirality occurs among the components of a system as a function of the interplay of kinetics and thermodynamics[29].

a



b

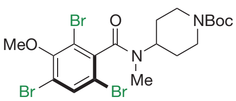
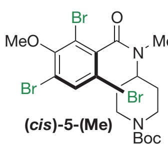
Compound	$\Delta G^{\ddagger}_{racemization}$ (experimental)	$\Delta G^{\ddagger}_{racemization}$ (theoretical)
 (*trans*)-5-(Me)	28.6	28.5
 (*cis*)-5-(Me)	27.8	28.2

Figure 6 | Energetic considerations and analysis of racemization dynamics. **a**, Computed geometries and modes of isomerization of **5-(Me)** with a concerted C-N/Ar-CO rotation and independent C-N rotation leading to racemization at high temperatures. **b**, Experimentally and theoretically calculated barriers to racemization of atropisomeric benzamides (kcal mol^{-1}). The experimental racemization data were obtained by heating samples and plotting ln(1/e.e.) versus time, where e.e. is enantiomeric excess. Individual *cis* and *trans* values were extracted using K_{eq} and the observed rate constant of racemization $k_{observed}$ of the isomeric mixture. The theoretical racemization data were computed using M062X/6-311++G(2d,3p)//B3LYP/6-31+G(d,p). Individual *cis* and *trans* values were determined from the computed free-energy difference. See Supplementary Information sections IX and XIV for further details.

METHODS SUMMARY

Dibromodimethylhydantoin (DBDMH, 85.8 mg, 0.30 mmol) was added to a 0.02 M solution of the aromatic amide **4** (66.9 mg, 0.20 mmol) and catalyst **1** (ref. 10) (11.5 mg, 0.02 mmol) in $CHCl_3$ (10 ml) at various temperatures (0 °C, −40 °C or −55 °C). The reaction was allowed to stir overnight (15–22 h). The reaction was then quenched with a 1.5 M solution of butyl vinyl ether in methanol (MeOH, 0.5 ml). The time of quench was recorded for amide equilibration calculations and defined as time zero (see Supplementary Information section VIII). For ease of chiral HPLC development, the phenol was subsequently protected as the methyl ether. The reaction was allowed to warm to room temperature and additional methanol was added (2.0 ml, ~5:1 by volume CH_2Cl_2:MeOH) followed by 2.0 M trimethylsilyldiazomethane in hexanes (0.4 ml). The methylation was quenched with silica gel upon completion (15–30 min), filtered and concentrated under reduced pressure. Flash chromatography of the crude residue with hexanes/EtOAc afforded products (as a *cis*–*trans* mixture or independently isolated isomers). This material (**5-Me**) was dissolved in a 10% *i*PrOH/hexanes mixture (0.1 M) and allowed to equilibrate, in solution, at room temperature over a prolonged time course to determine the experimental barriers to amide isomerization. Barriers to racemization of **5-Me** were determined experimentally by heating in toluene at 70 °C (0.6 M). Isomeric ratios were determined by chiral HPLC using a Chiralcel OD-H column, with a flow rate of 0.75 ml min^{-1}, in 95:5 hexanes:ethanol and retention times as follows: $T^{retention}_{(R,\,trans)} = 14.0$ min, $T^{retention}_{(S,\,trans)} = 15.3$ min, $T^{retention}_{(S,\,cis)} = 20.6$ min, $T^{retention}_{(R,\,cis)} = 23.9$ min.

Received 21 November 2013; accepted 26 February 2014.

Published online 20 April 2014.

1. Oki, M. *Topics in Stereochemistry* Vol. 1 *Atropisomerism* (Wiley Interscience, 1983).
2. Clayden, J., Moran, W. J., Edwards, P. J. & LaPlante, S. R. The challenge of atropisomerism in drug discovery. *Angew. Chem. Int. Edn* **48,** 6398–6401 (2009).
3. Feringa, B. L. The art of building small: from molecular switches to molecular motors. *J. Org. Chem.* **72,** 6635–6652 (2007).
4. Eliel, E. L. & Wilen, S. *Stereochemistry of Organic Compounds* 425 (Wiley Interscience, 1994).

5. Reist, M., Testa, B., Carrupt, P.-A., Jung, M. & Schurig, V. Racemization, diastereomerization, and epimerization: their meaning and pharmacological significance. *Chirality* **7,** 396–400 (1995).
6. Reichert, S. & Breit, B. Development of an axial chirality switch. *Org. Lett.* **9,** 899–902 (2007).
7. Wang, J. & Feringa, B. L. Dynamic control of chiral space in a catalytic asymmetric reaction using a molecular motor. *Science* **331,** 1429–1432 (2011).
8. Kelly, T. R., De Silva, H. & Silva, R. A. Unidirectional rotary motion in a molecular system. *Nature* **401,** 150–152 (1999).
9. LaPlante, S. R. *et al.* Assessing atropisomer axial chirality in drug discovery and development. *J. Med. Chem.* **54,** 7005–7022 (2011).
10. Barrett, K. T. & Miller, S. J. Enantioselective synthesis of atropisomeric benzamides through peptide-catalyzed bromination. *J. Am. Chem. Soc.* **135,** 2963–2966 (2013).
11. Keith, J. M., Larrow, J. F. & Jacobsen, E. N. Practical considerations in kinetic resolution reactions. *Adv. Synth. Catal.* **343,** 5–26 (2001).
12. Cox, C. & Lectka, T. Synthetic catalysis of amide isomerization. *Acc. Chem. Res.* **33,** 849–858 (2000).
13. Mannschreck, A., Mattheus, A. & Rissmann, G. Comparision of kinetic results obtained by NMR line shape and equilibration methods. *J. Mol. Spectrosc.* **23,** 15–31 (1967).
14. Ahmed, A. *et al.* Barriers to rotation about the chiral axis of tertiary aromatic amides. *Tetrahedron* **54,** 13277–13294 (1998).
15. Iwamura, H. & Mislow, K. Stereochemical consequences of dynamic gearing. *Acc. Chem. Res.* **21,** 175–182 (1988).
16. Bringmann, G. *et al.* Atroposelective synthesis chiral biaryl compounds. *Angew. Chem. Int. Edn* **44,** 5384–5427 (2005).
17. Erol, S. & Dogan, I. Determination of barriers to rotation of axially chiral 5-methyl-2-(*o*-aryl)imino-3-(*o*-aryl)thiazolidine-4-ones. *Chirality* **24,** 493–498 (2012).
18. Chupp, J. P. & Olin, J. F. Chemical and physical properties of some rotational isomers of α-haloacetanilides, a novel unreactive halogen system. *J. Org. Chem.* **32,** 2297–2303 (1967).
19. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120,** 215–241 (2008).
20. Campomanes, P., Menendez, M. I. & Sordo, T. L. A theoretical analysis of enantiomerization in aromatic amides. *J. Phys. Chem.* **106,** 2623–2628 (2002).
21. Bragg, R. A., Clayden, J., Morris, G. A. & Pink, J. H. Stereodynamics of bond rotation in tertiary aromatic amides. *Chemistry* **8,** 1279–1289 (2002).
22. Pirkle, W. H., Welch, C. J. & Zych, A. J. Chromatographic investigation of the slowly interconverting atropisomers of hindered naphthamides. *J. Chromatogr. A* **648,** 101–109 (1993).
23. Clayden, J. & Pink, J. H. Concerted rotation in a tertiary aromatic amide: toward a simple molecular gear. *Angew. Chem. Int. Edn* **37,** 1937–1939 (1998).
24. Cuyegkeng, M. A. & Mannschreck, A. Chromatographic separation of enantiomers and barriers to enantiomerization of axially chiral aromatic carboxamides. *Chem. Ber.* **120,** 803–809 (1987).
25. Lee, W. K., Park, Y. S. & Beak, P. Dynamic thermodynamic resolution: advantage by separation of equilibration and resolution. *Acc. Chem. Res.* **42,** 224–234 (2009).
26. Hirsch, R. & Hoffmann, R. W. A test on the configurational stability of chiral organolithium compounds based on kinetic resolution; scope and limitations. *Chem. Ber.* **125,** 975–982 (1992).
27. LaPlante, S. R., Edwards, P. J., Fader, L. D., Jakalian, A. & Hucke, O. Revealing atropisomer axial chirality in drug discovery. *ChemMedChem* **6,** 505–513 (2011).
28. LaPlante, S. R. *et al.* Enantiomeric atropisomers inhibit HCV polymerase and/or HIV matrix: characterizing hindered bond rotations and target selectivity. *J. Med. Chem.* **57,** 1944–1951 (2014).
29. Klussmann, M. *et al.* Thermodynamic control of asymmetric amplification in amino acid catalysis. *Nature* **441,** 621–623 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to the National Institute of General Medical Sciences of the NIH (GM-068649) for support. We also thank L. Guard for X-ray crystallography. A.J.M. was supported by the National Science Foundation Graduate Research Fellowship Program. All computational work was supported by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center, and by the National Science Foundation under grant number CNS 08-21132, which partially funded acquisition of the facilities.

Author Contributions S.J.M. and K.T.B. designed the project. K.T.B. performed the experiments. A.J.M. performed the theoretical calculations. All authors contributed to the analysis of data and composition of the manuscript.

Author Information Crystallographic data are deposited with the Cambridge Crystallographic Data Centre under the accession number CCDC 969575 (for **5-(X-ray)**). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.M. (scott.miller@yale.edu).

LETTER

doi:10.1038/nature13196

North Atlantic forcing of tropical Indian Ocean climate

Mahyar Mohtadi[1], Matthias Prange[1], Delia W. Oppo[2], Ricardo De Pol-Holz[3], Ute Merkel[1], Xiao Zhang[1], Stephan Steinke[1] & Andreas Lückge[4]

The response of the tropical climate in the Indian Ocean realm to abrupt climate change events in the North Atlantic Ocean is contentious. Repositioning of the intertropical convergence zone is thought to have been responsible for changes in tropical hydroclimate during North Atlantic cold spells[1–5], but the dearth of high-resolution records outside the monsoon realm in the Indian Ocean precludes a full understanding of this remote relationship and its underlying mechanisms. Here we show that slowdowns of the Atlantic meridional overturning circulation during Heinrich stadials and the Younger Dryas stadial affected the tropical Indian Ocean hydroclimate through changes to the Hadley circulation including a southward shift in the rising branch (the intertropical convergence zone) and an overall weakening over the southern Indian Ocean. Our results are based on new, high-resolution sea surface temperature and seawater oxygen isotope records of well-dated sedimentary archives from the tropical eastern Indian Ocean for the past 45,000 years, combined with climate model simulations of Atlantic circulation slowdown under Marine Isotope Stages 2 and 3 boundary conditions. Similar conditions in the east and west of the basin rule out a zonal dipole structure as the dominant forcing of the tropical Indian Ocean hydroclimate of millennial-scale events. Results from our simulations and proxy data suggest dry conditions in the northern Indian Ocean realm and wet and warm conditions in the southern realm during North Atlantic cold spells.

In the North Atlantic, the most recent glacial and deglacial periods are characterized by a series of abrupt and severe cold snaps of millennial duration associated with either iceberg instabilities and surges (Heinrich events) or freshwater input from the Arctic Ocean[6] (the Younger Dryas). These abrupt events are of particular interest because they were rapidly communicated through the ocean by a slowdown, or potentially a shutdown, of the Atlantic meridional overturning circulation[7] (AMOC) and through the atmospheric circulation[8] causing climate anomalies worldwide. Climate archives document a significant tropical hydrologic response to these events. Dry Younger Dryas and Heinrich stadials have been reported from various marine and terrestrial archives across the tropical Indian Ocean[4,9–14]. However, a few records suggest wet Younger Dryas or Heinrich stadials over northeast Australia[15], southern Indonesia[5,16] and southeast Africa[12,17].

Although there seems to be strong evidence that the intertropical convergence zone (ITCZ) moved southwards in the tropical Atlantic[2], a wide range of mechanisms have been offered to explain the connection between the cooling of the North Atlantic and tropical Indian Ocean hydroclimates: a weakening of the rainfall system in response to regional sea surface cooling[13,14]; and changes in the monsoon intensity[4,10,16] associated with a southward shift in the mean[1] or winter[4,5,15] position of the ITCZ or in the position of oceanic fronts[18]. However, there are no high-resolution records from the region of the tropical Indian Ocean unaffected by monsoon seasonality. Records from this region are needed to evaluate whether the eastern and western Indian Ocean experienced similar or contrasting rainfall anomalies during abrupt climate events. Furthermore, the Younger Dryas signal in many of the available records is ambiguous at best, and almost none of them include data associated with North Atlantic millennial events older than Heinrich stadial 1.

The effect of AMOC slowing on tropical climate can be studied with freshwater hosing experiments using coupled atmosphere–ocean general circulation models. Such numerical experiments consistently show a southward displacement of the ITCZ over the Atlantic Ocean in response to North Atlantic cooling induced by a perturbation of the AMOC[1,19,20]. However, these experiments provide ambiguous results

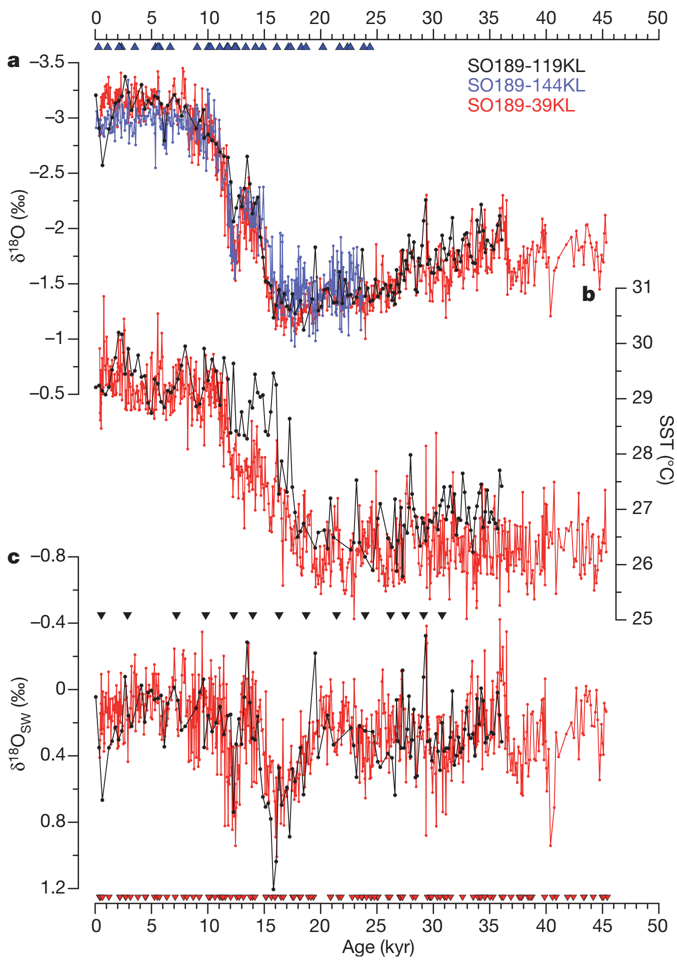


Figure 1 | Hydroclimate records from the eastern tropical Indian Ocean. Time series of $\delta^{18}O = (^{18}O/^{16}O)_{sample}/(^{18}O/^{16}O)_{standard} - 1$ (where the standard is PeeDee Belemnite (PDB)) (**a**), SST (**b**) and $\delta^{18}O_{SW}$ (seawater $\delta^{18}O$ with standard mean ocean water as the standard) (**c**). The chronology of each core is established independently by ^{14}C accelerator mass spectrometry dating (triangles; see Methods). There is a large similarity between all records during the time they overlap, with no glacial–interglacial difference in the sea-level-corrected $\delta^{18}O_{SW}$.

[1]MARUM-Center for Marine Environmental Sciences, University of Bremen, 28359 Bremen, Germany. [2]Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [3]Department of Oceanography, University of Concepción, Concepción, Chile. [4]Federal Institute for Geosciences and Natural Resources, 30655 Hannover, Germany.

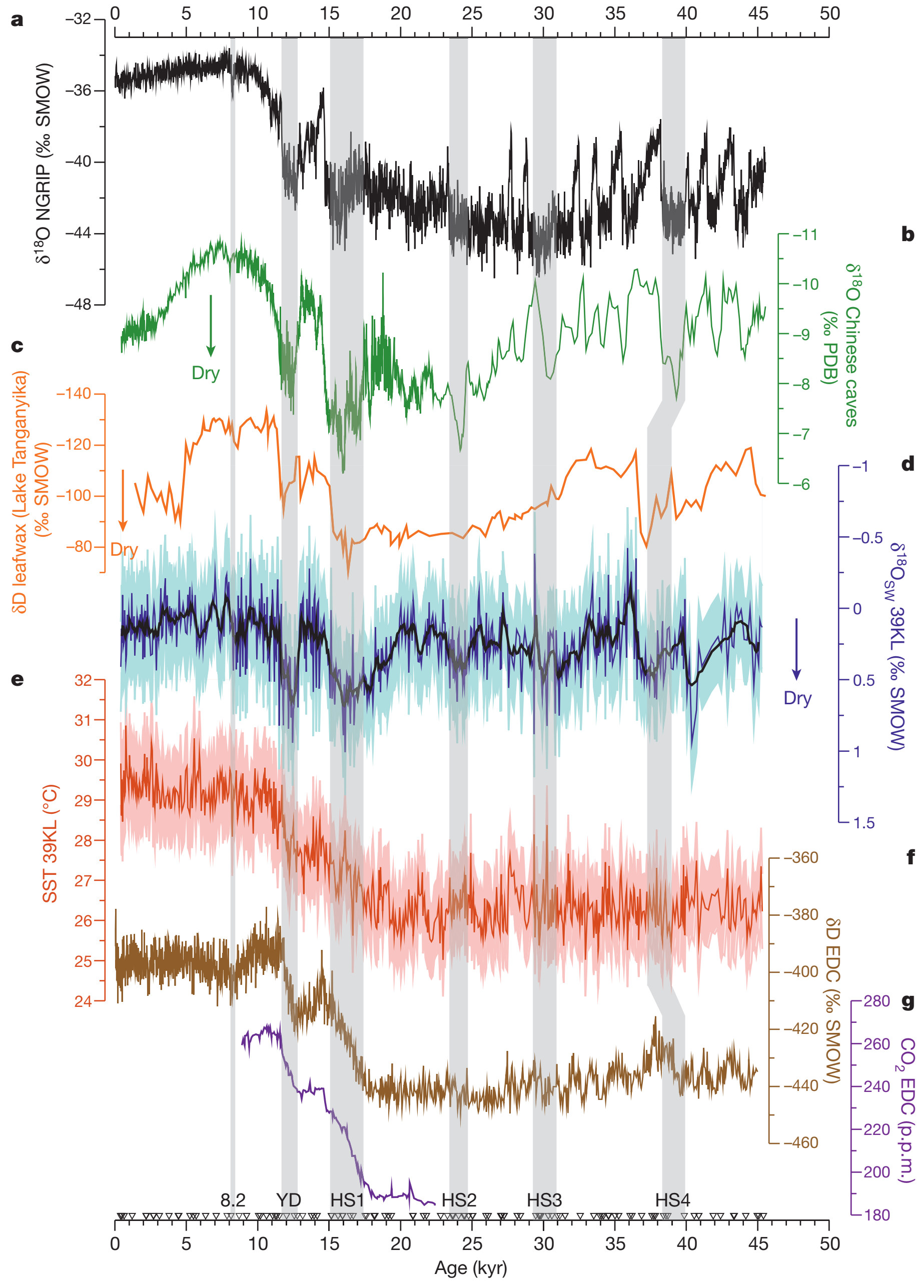


Figure 2 | Comparison of East Indian Ocean $\delta^{18}O_{SW}$ and SST data with other records of palaeoclimate. **a**, $\delta^{18}O$ data of Greenland ice core NGRIP[29]; **b**, stack $\delta^{18}O$ record of Chinese speleothems[30]; **c**, deuterium isotopes (δD) of leaf wax from Lake Tanganyika[14]; **d**, sea-level-corrected $\delta^{18}O_{SW}$ of core 39KL off western Sumatra (this study; black line represents a 7-point running average); **e**, SST reconstruction at site 39KL off western Sumatra (this study). Envelopes in **d** and **e** indicate 1σ errors (Methods). **f**, δD data from the EPICA Dome C ice core[22]. **g**, Deglacial CO_2 record from the EPICA Dome C ice core[22]. Grey bars indicate the 8.2 kyr event, the Younger Dryas (YD), and Heinrich stadials (HSs) 1 to 4 as recorded in North Atlantic deep-sea cores[23]. Triangles indicate the accelerator mass spectrometry ^{14}C age control points for core 39KL (this study).

for the tropical Indo-Pacific region, with the response and sensitivity of the region to North Atlantic freshwater perturbations being strongly model dependent[20]. Thus, palaeoclimate data are invaluable for identifying mechanisms of hydrologic change outside the Atlantic realm.

Here we present three high-resolution marine sedimentary records from the western coast of Sumatra spanning the past 45 kyr. The age models for these cores are well constrained and are based on a total of 162 radiocarbon dates (Methods and Supplementary Table 1). We measured shell Mg/Ca ratios (806 samples) and stable oxygen isotopes ($\delta^{18}O$; 1,191 samples) of the surface-dwelling planktonic foraminifera *Globigerinoides ruber sensu stricto* to reconstruct variations in seawater $\delta^{18}O$ ($\delta^{18}O_{SW}$; Methods). These records include the first high-resolution paired Mg/Ca–$\delta^{18}O_{SW}$ record (average sampling time, 60 yr) from the equatorial eastern Indian Ocean, which provides a measure of hydrological changes during the past 45 kyr (Fig. 1). Modern observations and instrumental records from the non-upwelling fore-arc basins off northern and central Sumatra, where our sites are located, suggest minor seasonal changes in precipitation, sea surface temperature (SST) and salinity compared with the upwelling region off south Java, and confirm that the study area lies outside the main Australasian monsoonal rainfall domains (Methods). Whereas records from other areas of the Indo-Pacific warm pool may be decoupled from local precipitation owing to oceanic advection[21], salinity at our core sites is primarily driven by local rainfall and runoff (Methods). We therefore infer that the observed changes in our proxy records reflect local precipitation and can be used to unravel the response of the tropical eastern Indian Ocean during the Younger Dryas and Heinrich events.

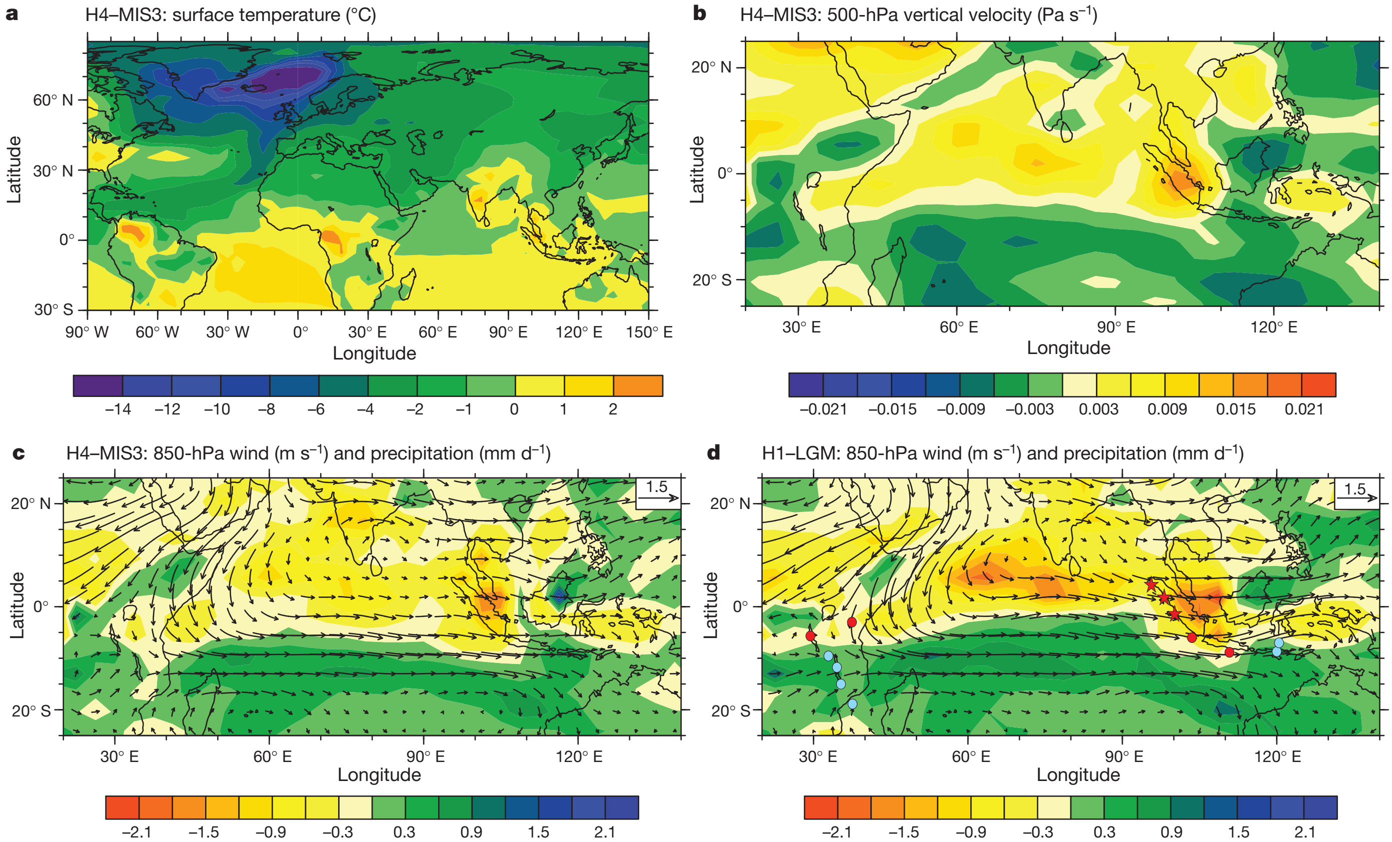


Figure 3 | Results from the CCSM3 simulations of Heinrich stadials 1 and 4. Plots show the difference between the Heinrich (hosing) experiments and the baseline simulations (MIS3 for H4 and LGM for H1; see Methods) as annual means. **a**, Surface temperature anomaly during H4, showing cooling over the Northern Hemisphere and warming in the Southern Hemisphere. **b**, Vertical velocity anomaly in the mid-level (500-hPa) atmosphere, with negative and positive values respectively indicating anomalously rising and sinking air during H4. The ascending branch of the annual mean Hadley circulation over the Indian Ocean is displaced southwards. **c**, 850-hPa wind (arrows) and precipitation (shading) anomalies during H4. Westerly low-level wind anomalies are associated with wetter southern Indian Ocean and drier equatorial and northern Indian Ocean. **d**, Same as **c** but for the simulation of H1. Note the similarity between **c** and **d**, and the competing influence of the Pacific Walker cell on the eastern part of the maritime continent. Red stars (this study) and dots[4,10,14,27] show sites indicating dry conditions during Heinrich stadial 1; blue dots show sites indicating wet Heinrich stadial 1 or Younger Dryas[5,12,16,17].

Reconstructed SST, $\delta^{18}O$ and $\delta^{18}O_{SW}$ from all sites show a remarkably consistent pattern during times when they overlap (Fig. 1), suggesting a coherent response from central to northern Sumatra. The SST records indicate that the deglacial warming of about 3 °C started 18,300 ± 330 yr ago (1σ; see Methods), synchronously (within dating uncertainties) with southern high-latitude climate change and an increase in atmospheric CO_2 concentration occurring during this period[22] (Fig. 2).

Although the sea-level-corrected $\delta^{18}O_{SW}$ records show no glacial–interglacial change, the values increase during the Northern Hemisphere climate cold spells, as recorded in Greenland ice cores. Our records suggest decreased precipitation in the tropical eastern Indian Ocean that is most prominent during Heinrich stadials 1 and 4 and the Younger Dryas (Fig. 2). The difference between the age of Heinrich stadial 4 as defined in our marine records from off Sumatra and that derived from Greenland ice cores or Chinese speleothems (Fig. 2) is within the average 2σ error, which is larger than 2 kyr in sediment older than 35 kyr (Methods). Thus, on the basis of the similar timing of younger millennial events in our $\delta^{18}O_{SW}$ record and North Atlantic cold events, we argue that the high-$\delta^{18}O_{SW}$ interval near the documented age of Heinrich event 4 defined by layers containing peaks of ice-rafted detritus in North Atlantic marine records[23] is also synchronous with that event.

To elucidate the dynamics of changes in the tropical Indian Ocean hydroclimate forced by the North Atlantic, we performed and analysed freshwater hosing experiments under Marine Isotope Stages (MISs) 2 and 3 boundary conditions using the comprehensive coupled climate model CCSM3. More precisely, baseline simulations of the Last Glacial Maximum (LGM; 21 kyr ago (MIS2)) and the time slice from 38 kyr ago (MIS3) were perturbed by 0.2 Sv (1 Sv = $10^6\ m^3\ s^{-1}$) freshwater injections into the northern North Atlantic for 500 yr, mimicking Heinrich events 1 and 4 (Methods), two of the most pronounced events in our records. North Atlantic water hosing results in a drastic slowdown of the AMOC by ~10 Sv in both glacial experiments (Methods) and, owing to the reduced northward Atlantic heat transport, a cooling of the surface of the North Atlantic (Fig. 3a). This cooling quickly propagates zonally throughout the Northern Hemisphere through atmospheric advection by the westerly winds and mixing[19]. The tropical mean circulation responds to the Northern Hemisphere cooling (and Southern Hemisphere warming) by generating anomalous energy transport from the southern hemisphere to the northern hemisphere. This energy transport is accomplished by a reorganization of the mean Hadley circulation involving an anomalous northward cross-equatorial flow in the upper branch accompanied by an anomalous southward flow in the lower branch[24]. Over the southern Indian Ocean, this results in a weakening of the annual mean Hadley circulation involving a southward shift of its rising branch, reflected by a north–south dipole structure in the mid-tropospheric vertical velocity anomaly (Fig. 3b). The reorganization of the Hadley circulation is associated with a westerly, low-level wind anomaly over the equatorial and south equatorial Indian Ocean (Fig. 3c; northwesterly at the surface (not shown)). This weakening of the southeast trade winds that cross the Equator reduces upwelling in the Arabian Sea (where there are weaker southwesterly winds), as also suggested by proxy records from this region[25]. Moreover, strengthening of the eastward winds over the Equator, a unique feature in the Indian Ocean responsible for warm equatorial waters, also contributes to the warming in the eastern tropical Indian Ocean.

The vertical velocity anomalies are associated with precipitation anomalies inducing drier conditions over the equatorial and north Indian Ocean and more humid conditions south of there (Fig. 3c). In particular, the model simulates a negative rainfall anomaly over Sumatra, whereas rainfall over southern Indonesia increases. The regional climatic response is qualitatively the same in both glacial hosing experiments (that is, under LGM and MIS3 boundary conditions), indicating that the same mechanisms may have been at work during all Heinrich stadials of the last glacial period (Fig. 3d and Methods). Despite the caveat of using a single model, the model results are consistent with our proxy data and also with Heinrich stadial 1 signals from other sites in the tropical Indian Ocean realm. Records suggesting humid conditions in Southern Indonesia and northern Australia during the Younger Dryas and Heinrich stadial 1[5,15,16] lie within the easternmost part of the anomalous Indian Ocean Hadley cell, which is confined by the ascending branch of the Pacific Walker cell (Fig. 3). Likewise, lake records from East Africa that suggest humid conditions during Heinrich stadial 1[12,17] lie beneath the ascending branch of the anomalous Hadley cell (Fig. 3). The remaining records of hydrological changes from the equatorial and northern Indian Ocean[26,27] (including our records) suggest drier conditions during Heinrich stadial 1, in agreement with our model simulation. As part of the Hadley cell reorganization during Heinrich stadials, the Indian summer monsoon weakens. This weakening has been recently attributed to anomalous SST forcing in the tropical Atlantic that affects the Indian monsoon via a tropical atmospheric pathway[3]. Our model results suggest a rapid response of the Indian monsoon to North Atlantic abrupt events through both a stationary Rossby wave-train teleconnection that originates in the northern North Atlantic and a tropical atmospheric pathway (Methods).

Our results suggest that the North Atlantic climate cold spells affect the hydrology of the tropical Indian Ocean through a reorganization of the Hadley circulation including a southward displacement of the ITCZ. This finding suggests that northern Indian Ocean cooling alone cannot explain tropical Indian Ocean hydrology during the Younger Dryas and Heinrich stadials. Indeed, our SST reconstructions show a slight warming during several Heinrich stadials, suggesting that at least in central-northern Sumatra, reduced rainfall was not a response to local SST cooling. Comparison of our data from the eastern tropical Indian Ocean with the Lake Tanganyika humidity record from East Africa[14] reveals a strong similarity ($r = 0.43$ with the range [0.11, 0.67], 95% confidence interval; Methods) with dry Younger Dryas and Heinrich stadials on both sides of the tropical Indian Ocean (Fig. 2), and suggests that zonally asymmetric changes within the Indian Ocean[14] were not the main control on Indian Ocean hydrology during these periods. Our results suggest that decreases in the monsoon intensity[28] stem from a reorganization of the Hadley cell over the Indian Ocean realm (Fig. 3), with heterogeneous mean annual and seasonal SST anomalies over the tropical Indian Ocean (Methods).

Integrated palaeoclimate data and model results from this study reveal that drastic changes in the tropical Indian Ocean climate appear as a robust response to the AMOC slowdown during Heinrich stadials and the Younger Dryas. This response involves similar mechanisms regardless of glacial background climate states. These climatic changes involve a reorganization of the Hadley circulation, with a southward shift of the ITCZ across the entire equatorial Indian Ocean. The resulting precipitation anomaly mirrors the meridional dipole-shaped rainfall anomaly in the tropical Atlantic region[2]. By contrast, our data suggesting little glacial–interglacial hydrologic change in the eastern Indian Ocean add to a growing body of evidence that the response of regional hydroclimate on millennial timescales was substantially greater than that on glacial–interglacial timescales[21,26].

METHODS SUMMARY

Piston cores SO189-119KL (3° 31′ N, 96° 19′ E; 780-cm core length, 808-m water depth), SO189-144KL (1° 9′ N, 98° 4′ E; 822-cm core length, 481-m water depth) and SO189-39KL (0° 47′ S, 99° 54′ E; 1,350-cm core length, 517-m water depth) were collected from the fore-arc basins off western Sumatra: the Simeulue basin (119KL), the Nias basin (144KL) and the northern Mentawai basin (39KL). Age models were based on linear interpolation between a total of 162 radiocarbon dates (Supplementary Table 1 and Extended Data Fig. 4). For $\delta^{18}O$ analysis, the isotopic composition of the carbonate sample was measured using a Finnigan MAT 251 mass spectrometer on the CO_2 gas evolved by treatment with phosphoric acid at a constant temperature of 75 °C. The Mg/Ca cleaning protocol consisted of five water and two methanol washes, two oxidation steps with 1% NaOH-buffered H_2O_2, and a weak acid leach with 0.001 M QD HNO_3. Samples were then dissolved into 0.075 M QD HNO_3 and centrifuged for 10 min at 6,000 r.p.m., transferred into test tubes and diluted. Mg/Ca ratios were measured using a Perkin Elmer Optima 3300 R inductively coupled plasma optical emission spectrophotometer for samples from core 199KL, and an Agilent Technologies 700 Series inductively coupled plasma optical emission spectrophotometer for samples from core 39KL. For details, see Methods and Extended Data. For the numerical experiments, the comprehensive global climate model CCSM3 was used. The baseline simulations for LGM and MIS3 comprise the orbital parameters, greenhouse gas concentrations, sea level and continental ice sheets at, respectively, 21 and 38 kyr ago. For the Heinrich stadial 1 and 4 experiments, the LGM and, respectively, MIS3 climate states were perturbed by a constant freshwater input of 0.2 Sv to the northern North Atlantic. The hosing experiments were integrated for 500 yr. For each experiment, the mean of the past 100 simulation years was used for analysis. For details, see Methods and Extended Data.

Data reported here are stored in the Pangaea database (www.pangaea.de).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 August 2013; accepted 28 February 2014.

1. Lewis, S. C. *et al.* High-resolution stalagmite reconstructions of Australian-Indonesian monsoon rainfall variability during Heinrich stadial 3 and Greenland interstadial 4. *Earth Planet. Sci. Lett.* **303,** 133–142 (2011).
2. Arbuszewski, J. A., deMenocal, P. B., Cleroux, C., Bradtmiller, L. & Mix, A. Meridional shifts of the Atlantic intertropical convergence zone since the Last Glacial Maximum. *Nature Geosci.* **6,** 959–962 (2013).
3. Marzin, C., Kallel, N., Kageyama, M., Duplessy, J.-C. & Braconnot, P. Glacial fluctuations of the Indian monsoon and their relationship with North Atlantic climate: new data and modelling experiments. *Clim. Past* **9,** 2135–2151 (2013).
4. Mohtadi, M. *et al.* Glacial to Holocene swings of the Australian-Indonesian monsoon. *Nature Geosci.* **4,** 540–544 (2011).
5. Muller, J., McManus, J. F., Oppo, D. W. & Francois, R. Strengthening of the Northeast Monsoon over the Flores Sea, Indonesia, at the time of Heinrich event 1. *Geology* **40,** 635–638 (2012).
6. Condron, A. & Winsor, P. Meltwater routing and the Younger Dryas. *Proc. Natl Acad. Sci. USA* **109,** 19928–19933 (2012).
7. McManus, J. F., Francois, R., Gherardi, J. M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428,** 834–837 (2004).
8. Moreno, P. I., Jacobson, G. L., Lowell, T. V. J. & Denton, G. H. Interhemispheric climate links revealed by a late-glacial cooling episode in southern Chile. *Nature* **409,** 804–808 (2001).
9. Deplazes, G. *et al.* Links between tropical rainfall and North Atlantic climate during the last glacial period. *Nature Geosci.* **6,** 213–217 (2013).
10. Mohtadi, M., Steinke, S., Lückge, A., Groeneveld, J. & Hathorne, E. C. Glacial to Holocene surface hydrography of the tropical eastern Indian Ocean. *Earth Planet. Sci. Lett.* **292,** 89–97 (2010).
11. Partin, J. W., Cobb, K. M., Adkins, J. F., Clark, B. & Fernandez, D. P. Millennial-scale trends in west Pacific warm pool hydrology since the Last Glacial Maximum. *Nature* **449,** 452–455 (2007).
12. Thomas, D. S. G., Burrough, S. L. & Parker, A. G. Extreme events as drivers of early human behaviour in Africa? The case for variability, not catastrophic drought. *J. Quaternary Sci.* **27,** 7–12 (2012).
13. Stager, J. C., Ryves, D. B., Chase, B. M. & Pausata, F. S. R. Catastrophic drought in the Afro-Asian monsoon region during Heinrich event 1. *Science* **331,** 1299–1302 (2011).
14. Tierney, J. E. *et al.* Northern hemisphere controls on tropical southeast African Climate during the past 60,000 years. *Science* **322,** 252–255 (2008).
15. Muller, J. *et al.* Possible evidence for wet Heinrich phases in tropical NE Australia: the Lynch's crater deposit. *Quat. Sci. Rev.* **27,** 468–475 (2008).
16. Griffiths, M. L. *et al.* Increasing Australian-Indonesian monsoon rainfall linked to early Holocene sea-level rise. *Nature Geosci.* **2,** 636–639 (2009).
17. Schefuß, E., Kuhlmann, H., Mollenhauer, G., Prange, M. & Pätzold, J. Forcing of wet phases in southeast Africa over the past 17,000 years. *Nature* **480,** 509–512 (2011).
18. De Deckker, P., Moros, M., Perner, K. & Jansen, E. Influence of the tropics and southern westerlies on glacial interhemispheric asymmetry. *Nature Geosci.* **5,** 266–269 (2012).
19. Clement, A. C. & Peterson, L. C. Mechanisms of abrupt climate change of the last glacial period. *Rev. Geophys.* **46,** RG4002 (2008).
20. Kageyama, M. *et al.* Climatic impacts of fresh water hosing under Last Glacial Maximum conditions: a multi-model study. *Clim. Past* **9,** 935–953 (2013).
21. Gibbons, F. T. *et al.* Deglacial $\delta^{18}O$ and hydrologic variability in the tropical Pacific and Indian Oceans. *Earth Planet. Sci. Lett.* **387,** 240–251 (2014).
22. Parrenin, F. *et al.* Synchronous change of atmospheric CO_2 and Antarctic temperature during the Last Deglacial Warming. *Science* **339,** 1060–1063 (2013).

23. Hemming, S. R. Heinrich events: massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. *Rev. Geophys.* **42,** RG1005 (2004).
24. Frierson, D. M. W. *et al.* Contribution of ocean overturning circulation to tropical rainfall peak in the Northern Hemisphere. *Nature Geosci.* **6,** 940–944 (2013).
25. Pourmand, A., Marcantonio, F. & Schulz, H. Variations in productivity and eolian fluxes in the northeastern Arabian Sea during the past 110 ka. *Earth Planet. Sci. Lett.* **221,** 39–54 (2004).
26. Carolin, S. A. *et al.* Varied response of Western Pacific hydrology to climate forcings over the Last Glacial Period. *Science* **340,** 1564–1566 (2013).
27. Verschuren, D. *et al.* Half-precessional dynamics of monsoon rainfall near the East African Equator. *Nature* **462,** 637–641 (2009).
28. Tierney, J. E. & deMenocal, P. B. Abrupt shifts in Horn of Africa hydroclimate since the Last Glacial Maximum. *Science* **342,** 843–846 (2013).
29. Svensson, A. *et al.* A 60000 year Greenland stratigraphic ice core chronology. *Clim. Past* **4,** 47–57 (2008).
30. Wang, Y. *et al.* Millennial- and orbital-scale changes in the East Asian monsoon over the past 224,000 years. *Nature* **451,** 1090–1093 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to K. Olafsdottir, M. Segl and B. Meyer-Schack for technical support. This study was funded by the German Bundesministerium für Bildung und Forschung (grant 03G0189A) and the Deutsche Forschungsgemeinschaft (DFG grants HE3412/15-1 and STE1044/4-1, and the DFG Research Centre/Cluster of Excellence 'The Ocean in the Earth System'). Climate model simulations were performed on the SGI Altix supercomputer of the Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen. D.W.O. is funded by the US NSF, R.D.P.-H. is supported by Chilean FONDAP 15110009/ICM Nucleus NC120066.

Author Contributions M.M., D.W.O. and A.L. designed the study. M.P., U.M. and X.Z. designed, performed and analysed the climate model experiments. M.M. and S.S. generated and analysed the proxy data. R.D.P.-H., M.M. and D.W.O. were responsible for the radiocarbon analyses. M.M. and M.P. wrote the manuscript; all authors discussed the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (mmohtadi@marum.de).

METHODS

Modern climate of the study area. At present the study area is characterized by only moderate monthly or seasonal changes in SST, air temperature, sea surface salinity (SSS) and precipitation (Extended Data Fig. 1). Mean annual SST averaged between 1854 and 2008 is about 29 °C with a small seasonal range of only 1.2 °C for the Simeulue basin, and 1.1 °C for the northern Mentawai and Nias basins (Extended Data Fig. 1, http://nomads.ncdc.noaa.gov/las/getUI.do). Likewise, 24-h air temperatures show little variability and are on average 26.7 °C with a small seasonal range of 2 °C near the core 119KL (station Sabang, averaged between 1976 and 1989; Extended Data Fig. 1) and one of 26.3 °C with a seasonal range of 0.9 °C near the core 39KL (station Padang, averaged between 1850 and 1989; http://climexp.knmi.nl; Extended Data Fig. 1). Generally, SST and air temperatures are highest during boreal spring because that is when near-surface winds are weakest, and are lowest during boreal fall owing to increased cloudiness and rainfall. Mean annual SSS is around 33.5 p.s.u. and varies seasonally within only ±0.2 p.s.u. (ref. 31). Instrumental records of precipitation between 1879 and 1989 show higher average monthly rainfall over central-western Sumatra (~360 mm per month) compared with northwestern Sumatra (~135 mm per month) (http://climexp.knmi.nl; Extended Data Fig. 1). Although there are two precipitation maxima in the study area during spring and fall related to the seasonal migration of the ITCZ[32], the contribution of each season to the total amount of rainfall does not vary considerably (~25% ± 5%; Extended Data Fig. 1).

These findings corroborate previous studies that exclude northern and western Sumatra from the Australasian monsoonal rainfall domains (see, for example, refs 32, 33). Observation and model studies suggest that there is no significant correlation between SST and rainfall variability in this region, between rainfall and ENSO[32], or between SST and ENSO[34]. In summary, the study area seems ideal to study the evolution of the hydrological changes in the tropical eastern Indian Ocean without significant biases introduced by variations in seasonal or interannual climate phenomena such as monsoon or ENSO.

The oceanic surface currents in the tropical eastern Indian Ocean flow according to the seasonally reversing monsoon winds. During boreal summer, a small branch of the northward flowing southwest monsoon current flows eastwards and joins the north equatorial counter current. This surface current is deflected southwards off the Sumatran coast and meets the south Java current off southwestern Sumatra (Extended Data Fig. 2). During summer, about 6 Sv of surface water flows southwards with the cross-equatorial meridional Ekman transport (MET), involving the export of low-salinity waters from the Bay of Bengal[35–37]. During boreal winter, the direction of currents in the tropical eastern Indian Ocean is reversed (Extended Data Fig. 2). Westwards flowing surface waters join the northeast monsoon current south of Sri Lanka and are affected by the northwards-directed MET[35,37]. Despite the seasonality in the currents, it seems that SST and SSS off northwestern Sumatra are not considerably affected by seasonally reversing surface current direction in the eastern Indian Ocean[37] or the strong salinity changes in the Bay of Bengal (see, for example, ref. 38; see also the next section).

Present and past control of $\delta^{18}O_{SW}$. The cores presented in this study have been collected close to the Sumatran coast that is separated from the open ocean by the Simeulue, Nias and Mentawai islands. During the last glacial period and Heinrich stadials, when sea level was up to 130 m lower than today, these fore-arc basins were more separated from the open ocean owing to their shallow sills of mostly <130 m (Extended Data Fig. 3). For the present, observation and model results suggest that "to the west of Sumatra, in particular, large rainfall and runoff persist year-round with enhanced surface stratification"[39] and that salinity off western Sumatra is strongly controlled by precipitation[40–42]. Western Sumatra receives an annual freshwater flux of ~1,800 mm yr^{-1}, where "a local maximum in P-E is present during all months, indicating that rainfall contributes to the existence of salinity stratified surface layer in this region"[42]. This overriding control of $\delta^{18}O_{SW}$ by local rainfall amount is also mirrored in our CTD data from two expeditions in 2005 and 2006 in this region: surface salinity at the core sites ranges between 32.5 p.s.u. (SO189-120MS at site 119KL and SO189-44MS at site 144KL) and 33 p.s.u. (SO189-40MS at site 39KL), whereas surface salinity west of the Mentawai Islands is as high as 34 p.s.u. (GeoB 10013-1 and 10018-1). This pattern supports previous findings that the low-salinity tongue off western Sumatra (and thus, $\delta^{18}O_{SW}$), particularly in the fore-arc basins of Mentawai, Nias and Simeulue, "is linked to the freshwater input in the near-surface layer from rainfall and river runoff"[43] rather than ocean advection. Sprintall *et al.*[44] and Janowiak and Xie[45] also stated that the freshwater pool located on the equator off the west Sumatra coast is related to a regional maximum in precipitation, as well as contribution from river runoff. We expect that these fore-arc basins were even more detached from large-scale surface circulation in the Indian Ocean during the last glacial period, when sea level was mostly below their sill depth.

In addition, a stalagmite record from Borneo[26] shows increased $\delta^{18}O$ values (of rainfall) during Heinrich stadials, in line with increased $\delta^{18}O_{SW}$ values in marine records from the Sulu Sea[46], the Lombok basin[47] and the Timor Sea[48] that lack any surface ocean connection to our study area. It is hard to explain such a consistent pattern in terrestrial and marine archives without involving changes in regional convective activity.

Material and methods. Piston cores SO189-119KL (3° 31′ N, 96° 19′ E; 780-cm core length, 808-m water depth), SO189-144KL (1° 09′ N, 98° 04 E; 822-cm core length, 481-m water depth) and SO189-39KL (0° 47′ S, 99° 54′ E; 1,350-cm core length, 517-m water depth) were collected from the upper continental margin in the Simeulue basin off northwestern Sumatra (119KL), from the Nias basin off northwestern Sumatra (144KL) and in the northern Mentawai basin off western Sumatra (39KL) during the RV *Sonne* cruise SO-189[49]. The cores were sampled at 5-cm (119KL) and 2-cm (144KL, 39KL) steps. Core-top studies from the Simeulue basin, the Nias basin and the northern Mentawai basin suggest that these basins lack a calcite lysocline[50,51]. The presence of well-preserved pteropods throughout the cores, along with the SEM observation of planktonic foraminifera in the corresponding core-tops, further supports the inference that selective calcite dissolution is, and has been, negligible at these sites.

Age models. The age model of SO189-119KL is based on 14 accelerator mass spectrometry (AMS) ^{14}C dates and linear interpolation between them (Supplementary Table 1 and Extended Data Fig. 4). AMS ^{14}C dating was performed on mixed planktonic foraminifera at the National Ocean Sciences Accelerator Mass Spectrometry Facility (NOSAMS) in Woods Hole, USA. The age model of SO189-39KL is based on linear interpolation between 112 AMS ^{14}C dates performed on mixed planktonic foraminifera at NOSAMS (16 dates), on *Globigerinoides ruber* (41 dates), *Globigerinoides sacculifer* without the final sac-like chamber (45 dates), and on mixed *G. ruber* and *G. sacculifer* (10 dates) at the Keck Carbon Cycle Accelerator Mass Spectrometry Facility at the University of California in Irvine, USA (Supplementary Table 1 and Extended Data Fig. 4). The age model of SO189-144KL is based on linear interpolation between 36 AMS ^{14}C dates performed on mixed planktonic foraminifera at NOSAMS (26 dates), on *G. ruber* (4 dates), on *G. sacculifer* without the final sac-like chamber (4 dates), on mixed *G. ruber* and *G. sacculifer* (1 date), and on mixed planktonic foraminifera (1 date) at the Keck Carbon Cycle Accelerator Mass Spectrometry Facility at the University of California in Irvine, USA (Supplementary Table 1 and Extended Data Fig. 4). All ages were corrected for ^{13}C, and ^{14}C ages were converted to calendar years using the CALIB 7.0-MARINE 13[52] program without any local offset. This assumption is based on studies from the Andaman Islands[53] and the Nicobar Islands[54] showing insignificant deviations from the global reservoir effect (ΔR values of 11 yr and 17 yr, respectively) in this region.

The correlation between the AMS ^{14}C dates in core 39KL shows statistically indistinguishable differences between the ages of monospecies and mixed species samples from the same core depths (Supplementary Fig. 1). We therefore included all the AMS ^{14}C dates in the age model and calculated an average age for each depth. In core 144KL, mixed planktonic radiocarbon ages are on average about 200 yr older than the ages of the surface-dwelling species (*G. ruber* and *G. sacculifer*) during the Holocene, and about 500 yr older during the last deglaciation (Supplementary Table 1). These age offsets are most probably due to the basin topography, which is characterized by relatively shallow sills that hamper a rigorous exchange between the subsurface waters of the Nias basin and the open ocean today. Accordingly, a lower sea-level stand during the last deglaciation increased the residence time of subsurface waters and the age offset between pure surface and mixed surface–subsurface radiocarbon dates to 500 yr. Therefore, the mixed planktonic ages were corrected by 200 yr (Holocene) and 500 yr (last deglaciation), respectively. Additional control of the calculated age model for core 144KL is provided by its stable oxygen isotope ($\delta^{18}O$) record, which perfectly matches the $\delta^{18}O$ records of cores 39KL and 119KL (Fig. 1 and main text). According to the age models, the average sedimentation rate (SR) at 119KL is about 24 cm kyr^{-1}, with a higher average SR during MIS3 (~29 cm kyr^{-1}) than during the last deglaciation (~20 cm kyr^{-1}) and the Holocene (~19 cm kyr^{-1}). The average SR at 144KL is about 43 cm kyr^{-1}, with a higher average SR during the LGM (~51 cm kyr^{-1}) and the last deglaciation (~49 cm kyr^{-1}) than during the Holocene (~31 cm kyr^{-1}). The average SR at 39KL is about 34 cm kyr^{-1}, with no considerable differences between MIS3 (~32 cm kyr^{-1}), the last deglaciation (~34 cm kyr^{-1}) and the Holocene (~37 cm kyr^{-1}).

Stable oxygen isotope ($\delta^{18}O$), Mg/Ca and $\delta^{18}O_{SW}$. In total, 147 (651) samples for $\delta^{18}O$ and 135 (671) samples for Mg/Ca were analysed in core 119KL (39KL). Additional 393 samples were analysed for $\delta^{18}O$ in core 144KL. $\delta^{18}O$ and Mg/Ca analyses were performed on about 30 tests of *G. ruber sensu stricto*[55] from the 250–355-μm size fraction. Previous sediment trap[56] and sediment surface[50,51] studies suggest that *G. ruber* tests record mean annual mixed-layer conditions in the tropical eastern Indian Ocean, and, in particular, mean annual surface conditions in the study area[51].

For $\delta^{18}O$ analysis, the isotopic composition of the carbonate sample was measured using a Finnigan MAT 251 mass spectrometer on the CO_2 gas evolved by treatment with phosphoric acid at a constant temperature of 75 °C. For all stable isotope measurements, a working standard was used, which has been calibrated against VPDB (Vienna Pee Dee Belemnite) by using the NBS 19 standard. Long-term analytical

standard deviation is about ±0.07‰ (Isotope Laboratory at Faculty of Geosciences, University of Bremen).

For Mg/Ca analysis, samples were cleaned by applying a slight modification of the method originally proposed in ref. 57, consisting of five water washes and two methanol washes followed by two oxidation steps with 1% NaOH-buffered H_2O_2 and then a weak acid leach with 0.001 M QD HNO_3. Samples were then dissolved into 0.075 M QD HNO_3 and centrifuged for 10 min at 6,000 r.p.m., transferred into test tubes and diluted. Mg/Ca ratios were measured using a Perkin Elmer Optima 3300 R inductively coupled plasma optical emission spectrophotometer (ICP-OES) equipped with an autosampler and a U-5000 AT ultrasonic nebulizer (Cetac Technologies) for samples from core 199KL, or an Agilent Technologies 700 Series ICP-OES with a CETAX ASX-520 autosampler for samples from core 39KL (both housed at the Faculty of Geosciences, University of Bremen). Mg/Ca values are reported as mmol mol^{-1}. The instrumental precision was determined using an external, in-house standard (Mg/Ca = 2.92 mmol mol^{-1}), which was run after every fifth sample. In core SO189-119KL, the relative standard deviations were 0.008 mmol mol^{-1} (0.26%) for the external standard and 0.07 mmol mol^{-1} (1.94%) for the ECRM 752-1 standard[58]. Replicate measurements on 30 samples revealed an average standard deviation of 0.15 mmol mol^{-1}. In core SO189-39KL, the relative standard deviations were 0.005 mmol mol^{-1} (0.16%) for the external standard and 0.06 mmol mol^{-1} (1.66%) for the ECRM 752-1 standard. Replicate measurements on 55 samples revealed an average standard deviation of 0.12 mmol mol^{-1}. Clay contamination and post-depositional Mn-rich carbonate/oxyhydroxide coatings could be excluded by measuring Fe/Ca, Mn/Ca and Al/Ca ratios (<0.1 mmol mol^{-1} for Mn/Ca and Fe/Ca; not detectable for Al/Ca). Mg/Ca ratios were converted to temperature following[59]

$$\mathrm{Mg/Ca}\,[\mathrm{mmol\,mol^{-1}}] = 0.38\mathrm{e}^{0.09T\,[^{\circ}\mathrm{C}]} \quad (1)$$

To calculate $\delta^{18}O_{SW}$ as a measure of hydrologic changes, we used the following δ^{18}O:temperature equation[60]:

$$T\,[^{\circ}\mathrm{C}] = 14.9 - 4.8(\delta^{18}\mathrm{O_{cc}} - \delta^{18}\mathrm{O_{SW}}) \quad (2)$$

where $\delta^{18}O_{cc}$ is the measured δ^{18}O of calcite and T is the Mg/Ca-derived temperature. The values were then converted to standard mean ocean water (SMOW) by adding 0.27‰ and corrected for sea-level changes proposed in ref. 61.

Error analysis for SST and $\delta^{18}O_{SW}$. The errors in $\delta^{18}O_{sw}$ and SST reconstructions are estimated by propagating the error introduced by the $\delta^{18}O_{cc}$ and Mg/Ca measurements (see previous section), the Mg/Ca:temperature calibration (equation (1)), the δ^{18}O:temperature equation (equation (2)) and the removal of the global ice volume[61] (±0.09‰). For the SST estimate, the following equation was used to propagate the errors by assuming no covariance among the errors[62]:

$$\sigma_T^2 = \left(\frac{\partial T}{\partial a}\sigma_a\right)^2 + \left(\frac{\partial T}{\partial b}\sigma_b\right)^2 + \left(\frac{\partial T}{\partial \mathrm{Mg/Ca}}\sigma_{\mathrm{Mg/Ca}}\right)^2$$

where

$$a = 0.090 \pm 0.003\ ^{\circ}\mathrm{C}^{-1}$$

$$b = 0.38 \pm 0.02\ \mathrm{mmol\,mol^{-1}}$$

$$\frac{\partial T}{\partial a} = -\frac{1}{a^2}\ln\left(\frac{\mathrm{Mg/Ca}}{b}\right)$$

$$\frac{\partial T}{\partial b} = -\frac{1}{ab}$$

and

$$\frac{\partial T}{\partial \mathrm{Mg/Ca}} = \frac{1}{a}\frac{1}{\mathrm{Mg/Ca}}$$

For the $\delta^{18}O_{sw}$ estimate, the following equation was used to propagate the errors by assuming no covariance among the errors[62]:

$$\sigma^2_{\delta^{18}\mathrm{O_{SW}}} = \left(\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial T}\sigma_T\right)^2 + \left(\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial a}\sigma_a\right)^2 + \left(\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial b}\sigma_b\right)^2 + \left(\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial \delta^{18}\mathrm{O_{cc}}}\sigma_{\delta^{18}\mathrm{O_{cc}}}\right)^2$$

where

$$a = 14.9 \pm 0.1\ ^{\circ}\mathrm{C}$$

$$b = -4.8 \pm 0.08\ ^{\circ}\mathrm{C}$$

$$\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial T} = -\frac{1}{b}$$

$$\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial a} = \frac{1}{b}$$

$$\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial b} = \frac{T}{b^2} - \frac{a}{b^2}$$

and

$$\frac{\partial \delta^{18}\mathrm{O_{SW}}}{\partial \delta^{18}\mathrm{O_{cc}}} = 1$$

The resulting errors are on average about 1 °C for SST and 0.3‰ for $\delta^{18}O_{SW}$ in cores 39KL and 119KL (Extended Data Fig. 5, Fig. 2 and main text). We note that the absolute SST values in 119KL are about 1 °C higher than in 39KL during the last deglaciation (Fig. 1 and main text). This difference is probably related to the more open Simeulue basin (site 119KL; see Extended Data Fig. 3). However, the calculated $\delta^{18}O_{SW}$ values for cores 119KL and 39KL are very similar, suggesting little influence of SST on $\delta^{18}O_{SW}$ and that the latter is mainly controlled by precipitation changes.

Estimating the timing of deglacial warming. To estimate the timing of deglacial warming from the Mg/Ca record of core 39KL, we applied a ramp-fitting method using RAMPFIT version 1.10 (http://www.manfredmudelsee.com/soft/rampfit)[63]. RAMPFIT has successfully been applied in previous studies for change-point detection in palaeoclimatic time series[64–67]. The simple ramp-function (three-phase) regression model assumes the existence of two distinct climate states, each of which is characterized by a constant parameter (here given by the SSTs of the cold glacial and the warm Holocene state), and a linear transition from one state to the other. Superimposed on this ramp function are short-term climate variations and noise from measurement uncertainties making visual determination of the two break points (that is, the start and end of the transition) difficult. RAMPFIT uses weighted least-squares regression to determine the amplitude of the transition, and a brute-force search (that is, a method that consists of systematically enumerating all possible candidates for the solution) for detecting start and end points. The time-dependent standard deviation provides the weights for the least-squares regression. Non-parametric stationary bootstrap re-sampling was used to estimate uncertainties in the timing of the transition, and age model uncertainties were not taken into account[63].

Because the SST time series of core 39KL shows a warming trend during the mid-to-late Holocene, data younger than 5 kyr were excluded from the ramp regression analysis. Taking the entire glacial period (that is, up to 45 kyr ago, the end of the SST record) into account, RAMPFIT estimates the onset of deglacial warming at 18.30 ± 0.33 kyr ago (1σ). This value is robust against the chosen interval length; for example, excluding MIS3 (that is, data older than 29 kyr) from the regression analysis leads to an estimated onset of deglacial warming at 18.29 ± 0.44 kyr ago (Extended Data Fig. 6). The break points between the deglacial transition and the early Holocene warm state were found to occur 9.62 ± 0.33 kyr ago and 9.75 ± 0.39 kyr ago for the 5–45-kyr and 5–29-kyr intervals, respectively (Extended Data Fig. 6). Mean Mg/Ca SSTs before and after the deglacial transition are 26.29 ± 0.03 and 29.18 ± 0.06 °C, respectively, when using the 5–45-kyr interval, and are 26.27 ± 0.06 and 29.16 ± 0.07 °C when using the 5–29 kyr interval.

Ramp-function regression has not been applied to the Mg/Ca record of core 119KL because its temporal resolution is too coarse. However, visual inspection suggests that deglacial warming started ~18 kyr ago, that is, synchronous with changes in atmospheric CO_2 and high-latitude temperature estimated from Antarctic ice cores[22].

Correlation between $\delta^{18}O_{SW}$ in 39KL and δD in Lake Tanganyika. The correlation between the $\delta^{18}O_{SW}$ record of SO189-39KL and the δD record of Lake Tanganyika[14] (Fig. 2 and main text) was tested by using the improved version of PEARSONT[68] (http://www.manfredmudelsee.com/soft/pearsont/index.htm). The software estimates the correlation coefficient (r) with accurate bootstrap confidence intervals by accounting for the autocorrelation (memory) of the data[68]. The two data sets were re-sampled at 0.3-kyr steps on the basis of their original age models to achieve the same timescale. The correlation coefficient and 95% confidence interval for the period 1.5–45 kyr ago is $r = 0.43$ [0.11, 0.67]. Because the confidence interval does not contain zero, the correlation between the two records is significant.

Numerical experiments. For the numerical experiments, the comprehensive global climate model CCSM3 (Community Climate System Model version 3, National Center for Atmospheric Research) was used. CCSM3 is a state-of-the-art, fully coupled model, composed of four separate components representing atmosphere, ocean, land and sea ice[69]. In our simulations, the resolution of the atmospheric component is 3.75° with 26 layers in the vertical, and the ocean has a nominal resolution of 3° with

equatorial grid refinement in the meridional direction (down to 0.9°) and 25 levels in the vertical[70]. The land model is defined on the same horizontal grid as the atmosphere and includes components for biogeophysics, biogeochemistry and the hydrologic cycle, as well as a dynamic global vegetation model[71,72]. To improve the simulation of the land surface hydrology and vegetation cover, new parameterizations for canopy interception and soil evaporation have been implemented into the land component[73]. Our experimental framework comprises four glacial simulations.

(1) A baseline simulation for the LGM following the guidelines of the Paleoclimate Modelling Intercomparison Project, Phase 2[74]. The boundary conditions for this simulation comprise the orbital parameters[75] and the greenhouse gas concentrations of CO_2, CH_4 and N_2O (185 p.p.m., 350 p.p.b. and 200 p.p.b., respectively) from 21 kyr ago. A sea-level lowering by 120 m has been taken into account by modifying the land–sea distribution, which, in particular, leads to closure of the Bering Strait. In addition, LGM continental ice sheets have been implemented on the basis of the ICE-5G data set[76]. The experimental set-up differs from the LGM run in ref. 77 only in the land surface component (see above). Our LGM simulation was integrated for 1,500 yr, starting from the quasi-equilibrated LGM run of ref. 77.

(2) A Heinrich stadial 1 analogue experiment in which the LGM climate state is perturbed by a constant freshwater input of 0.2 Sv to the northern North Atlantic, leading to a substantial weakening of the AMOC. The hosing experiment was integrated for 500 yr. Note that the applied freshwater forcing is in the range of published estimates for meltwater input during Heinrich events[23].

(3) A MIS3 baseline simulation in which the LGM boundary conditions (see above) were modified by applying the orbital forcing of 38 kyr ago and the corresponding greenhouse gas concentrations of CO_2, CH_4 and N_2O (215 p.p.m., 501 p.p.b. and 234 p.p.b., respectively; refs 78–80). In addition, the 38-kyr-ago ICE-5G continental ice-sheet distribution was implemented[76]. The MIS3 baseline simulation was initialized with the final state of our LGM simulation and integrated for another 2,100 yr.

(4) A Heinrich stadial 4 simulation in which the MIS3 (38 kyr ago) baseline climate is perturbed by a constant 0.2-Sv freshwater influx analogous to that in the Heinrich stadial 1 experiment. The Heinrich stadial 4 hosing experiment started at year 1,600 of our MIS3 baseline run and was integrated for 500 yr.

In all simulations, ozone and aerosol distributions were kept at pre-industrial levels[81]. For each experiment, the mean of the last 100 simulation years was used for analysis.

Extended Data Fig. 7 shows the Atlantic meridional overturning stream functions for the four different glacial climate states along with a pre-industrial reference run. The LGM North Atlantic overturning is slightly (1–2 Sv) stronger than the pre-industrial circulation, but the southward flow of North Atlantic Deep Water takes place at shallower depth. The AMOC is similarly strong in the MIS3 baseline run. On freshwater input to the North Atlantic, the AMOC weakens rapidly under both LGM and MIS3 boundary conditions.

Extended Data Fig. 8 shows the annual mean climatic response to AMOC slowdown under LGM boundary conditions (Heinrich stadial 1 experiment) for large-scale surface temperature and mid-tropospheric vertical velocity in the Indian Ocean region. Comparing the results of the Heinrich stadial 1 experiment with the Heinrich stadial 4 run (Fig. 3 and main text) suggests that the same mechanisms are at work, involving a southward shift of the ascending branch of the annual mean Hadley circulation. For the same region, the mid-tropospheric vertical velocity is shown in Supplementary Fig. 2 for both the LGM and MIS3 baseline simulations as a reference. Rising air over the equatorial region with subsidence to the north and south reveals the regional annual mean Hadley circulation in both climate states.

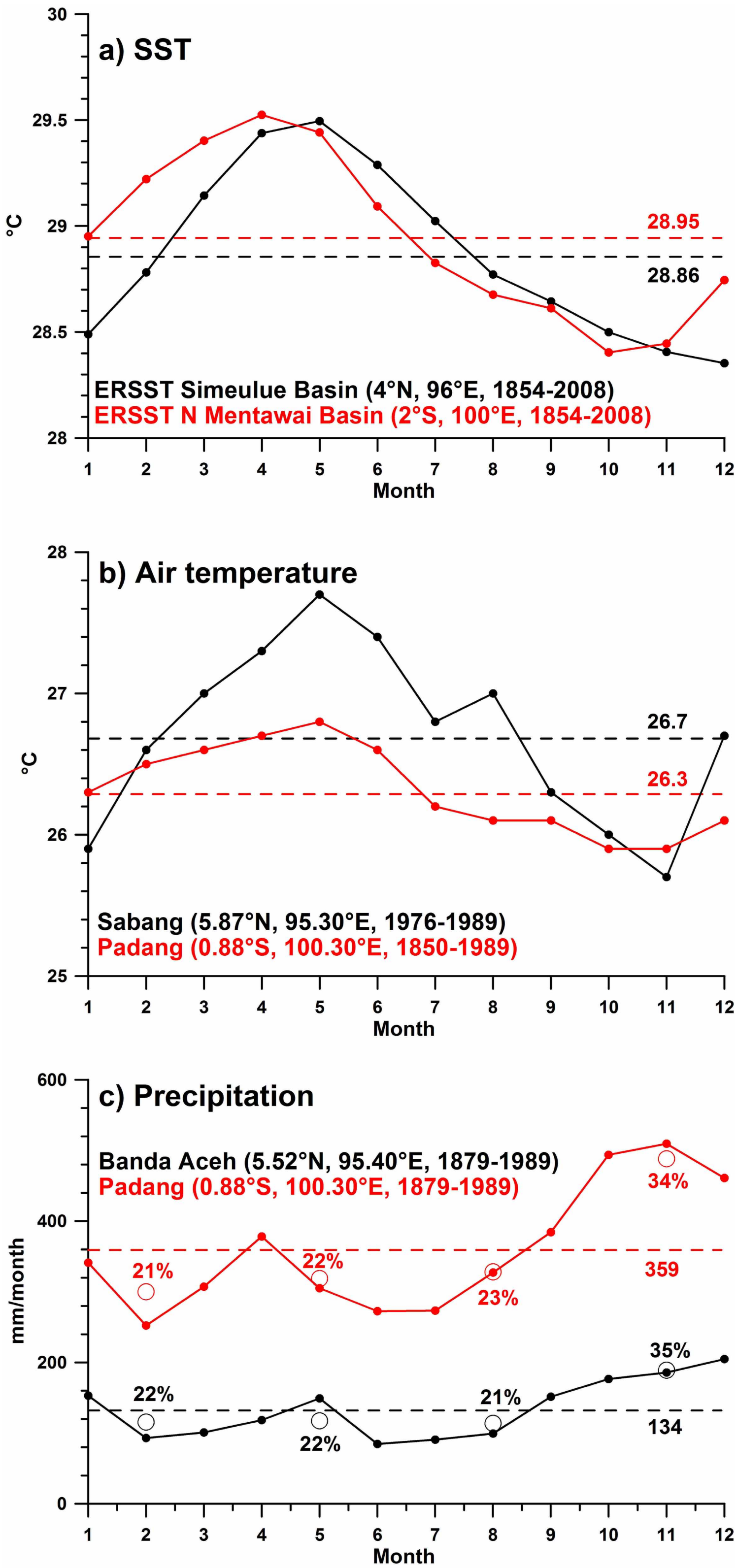
Indian summer monsoon weakening and atmospheric Atlantic Ocean/India teleconnection. As part of the Hadley cell reorganization during Heinrich stadials, the Indian summer monsoon weakens (Extended Data Fig. 9). The model results suggest a rapid response of the Indian monsoon to North Atlantic cooling through both a stationary Rossby wave train teleconnection that originates in the northern North Atlantic and a tropical atmospheric pathway.

The 200-hPa wind field anomaly (Heinrich stadial 4 hosing experiment minus MIS3 baseline run) in summer exhibits an arch-shaped wave train that propagates southeastwards from the northern North Atlantic region of perturbation to west-central Asia via east-central Europe, and farther, zonally across Asia towards the Pacific Ocean along the westerly jet (Extended Data Fig. 10). Except north and northwest of India, this wave train has an equivalent barotropic structure (not shown). From analysing modern observational data, a similar wave train has been identified as part of a circumglobal teleconnection pattern in the summertime mid-latitude circulation of the Northern Hemisphere, which significantly correlates with Indian monsoonal rainfall at the interannual timescale[82]. It has been suggested that the upper-level circulation anomaly affects Indian rainfall by changing the intensity of the monsoonal easterly vertical shear and, hence, the monsoon dynamics[82,83] (and moist dynamic instability). A westerly flow anomaly that stretches from North Africa to western India via the Arabian Sea adds to the upper-level circulation anomaly in the Indian monsoon region (Extended Data Fig. 10). A similar circulation anomaly over Africa and the Arabian Sea was recently found in freshwater hosing experiments[3] and, by means of atmosphere-only sensitivity experiments, could be attributed to anomalous SST forcing in the tropical Atlantic ('tropical pathway').

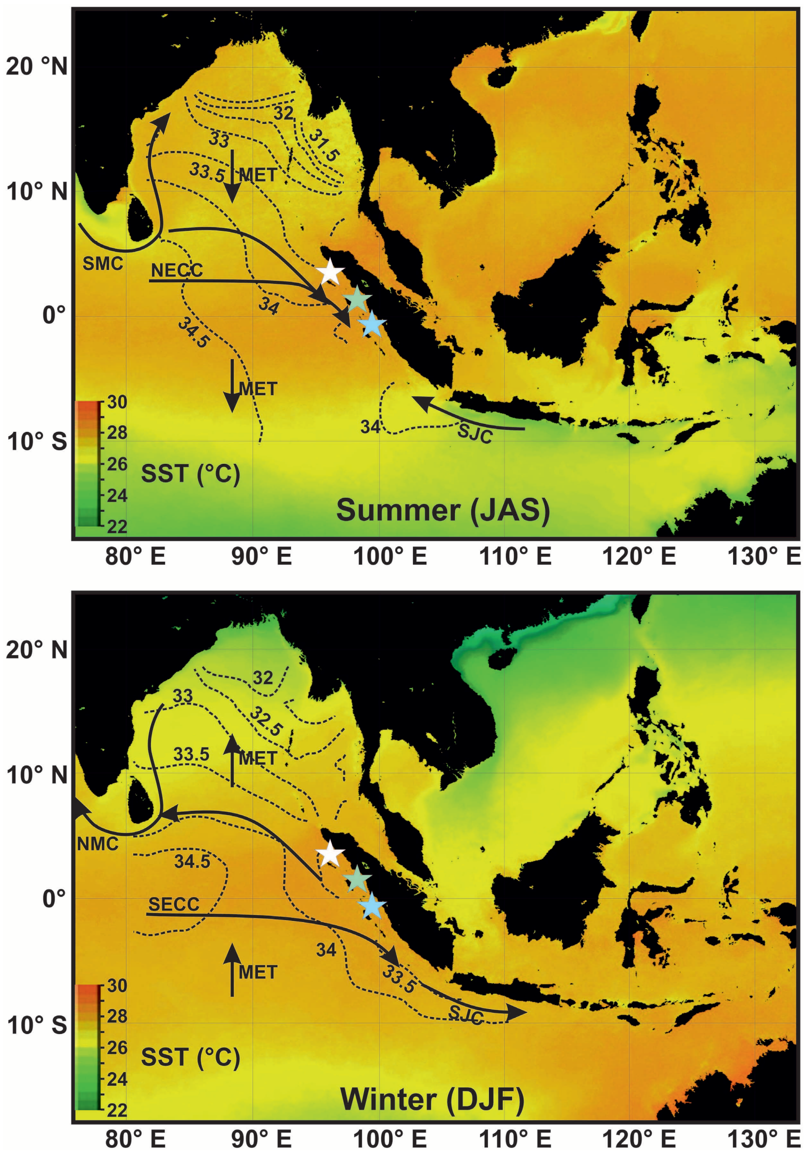
Unrelated to the Indian summer monsoon, an equivalent barotropic wave train propagating southeastwards from the northern North Atlantic source region is also found in the Heinrich stadial winter circulation anomaly (Extended Data Fig. 10). Compared with the summer season, the wave path is shifted towards the Equator following the seasonal jet stream displacement, crossing the Mediterranean Sea and the Arabian Peninsula until it reaches India, where an anticyclonic circulation anomaly develops in the upper troposphere. We surmise that this anomalous anticyclone promotes the upper-level southeasterly cross-equatorial flow over the Indian Ocean (Extended Data Fig. 10), which contributes to anomalous energy transport from the Southern Hemisphere to the anomalously cold Northern Hemisphere (compare with refs 24, 84). The increased energy demand of the Northern Hemisphere during Heinrich stadials is particularly large in the boreal winter season, when the Northern Hemisphere cooling is strongest (Supplementary Fig. 3), mainly owing to sea-ice effects (increased surface albedo and reduced ocean–atmosphere heat flux in the northern North Atlantic). We note that atmospheric circulation and precipitation response patterns are qualitatively the same in the Heinrich stadial 1 and Heinrich stadial 4 experiments (not shown).

31. Antonov, J. I., Locarnini, R. A., Boyer, T. P., Mishonov, A. V. & Garcia, H. E. *World Ocean Atlas 2005 Volume 2: Salinity* 182 (US Government Printing Office, 2006).
32. Aldrian, E. & Susanto, R. D. Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *Int. J. Climatol.* **23,** 1435–1452 (2003).
33. Wang, B. & Ding, Q. Global monsoon: dominant mode of annual variation in the tropics. *Dyn. Atmos. Oceans* **44,** 165–183 (2008).
34. Yoo, S.-H., Yang, S. & Ho, C.-H. Variability of the Indian Ocean sea surface temperature and its impacts on Asian-Australian monsoon climate. *J. Geophys. Res.* **111,** D03108 (2006).
35. Schott, F. A. & McCreary, J. P. The monsoon circulation of the Indian Ocean. *Prog. Oceanogr.* **51,** 1–123 (2001).
36. Sengupta, D., Raj, G. N. B. & Shenoi, S. S. C. Surface freshwater from Bay of Bengal runoff and Indonesian Throughflow in the tropical Indian Ocean. *Geophys. Res. Lett.* **33,** L22609 (2006).
37. Schott, F. A., Xie, S.-P. & McCreary, J. P. Jr. Indian Ocean circulation and climate variability. *Rev. Geophys.* **47,** RG1002 (2009).
38. Vinayachandran, P. N., Murty, V. S. N. & Ramesh Babu, V. Observations of barrier layer formation in the Bay of Bengal during summer monsoon. *J. Geophys. Res.* **107,** 8018 (2002).
39. Qu, T., Du, Y., Strachan, J., Meyers, G. & Slingo, J. M. Sea surface temperature and its variability in the Indonesian region. *Oceanography* **18,** 50–61 (2005).
40. Du, Y., Qu, T., Meyers, G., Masumoto, Y. & Sasaki, H. Seasonal heat budget in the mixed layer of the southeastern tropical Indian Ocean in a high-resolution ocean general circulation model. *J. Geophys. Res.* **110,** C04012 (2005).
41. Qu, T. & Meyers, G. Seasonal variation of barrier layer in the southeastern tropical Indian Ocean. *J. Geophys. Res.* **110,** C11003 (2005).
42. Sprintall, J. & Tomczak, M. Evidence of the barrier layer in the surface layer of the tropics. *J. Geophys. Res.* **97,** 7305–7316 (1992).
43. Qiu, Y., Cai, W., Li, L. & Guo, X. Argo profiles variability of barrier layer in the tropical Indian Ocean and its relationship with the Indian Ocean Dipole. *Geophys. Res. Lett.* **39,** L08605 (2012).
44. Sprintall, J., Potemra, J. T., Hautala, S. L., Bray, N. A. & Pandoe, W. W. Temperature and salinity variability in the exit passages of the Indonesian Throughflow. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **50,** 2183–2204 (2003).
45. Janowiak, J. E. & Xie, P. CAMS-OPI: a global satelliterain gauge merged product for real-time precipitation monitoring applications. *J. Clim.* **12,** 3335–3342 (1999).
46. Rosenthal, Y., Oppo, D. W. & Linsley, B. K. The amplitude and phasing of climate change during the last deglaciation in the Sulu Sea, western equatorial Pacific. *Geophys. Res. Lett.* **30,** 1428 (2003).
47. Levi, C. *et al.* Low-latitude hydrological cycle and rapid climate changes during the last deglaciation. *Geochem. Geophys. Geosyst.* **8,** Q05N12 (2007).
48. Zuraida, R. *et al.* Evidence for Indonesian Throughflow slowdown during Heinrich events 3-5. *Paleoceanography* **24,** PA2205 (2009).
49. Wiedicke-Hombach, M. *et al. SUMATRA - The Hydrocarbon System of the Sumatra Forearc*. Vol. Archive No. 0126492 (Federal Institute for Geosciences and Natural Resources Hannover, 2007).
50. Mohtadi, M. *et al.* Modern environmental conditions recorded in surface sediment samples off W and SW Indonesia: planktonic foraminifera and biogenic compounds analyses. *Mar. Micropaleontol.* **65,** 96–112 (2007).
51. Mohtadi, M. *et al.* Reconstructing the thermal structure of the upper ocean: Insights from planktic foraminifera shell chemistry and alkenones in modern sediments of the tropical eastern Indian Ocean. *Paleoceanography* **26,** PA3219 (2011).
52. Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55,** 1869–1887 (2013).
53. Dutta, K., Bhushan, K. & Somayajulu, B. L. K. ΔR correction values for the northern Indian Ocean. *Radiocarbon* **43,** 483–488 (2001).
54. Southon, J., Kashgarian, M., Fontugne, M., Metivier, B. & Yim, W. W.-S. Marine reservoir corrections for the Indian Ocean and southeast Asia. *Radiocarbon* **44,** 167–180 (2002).

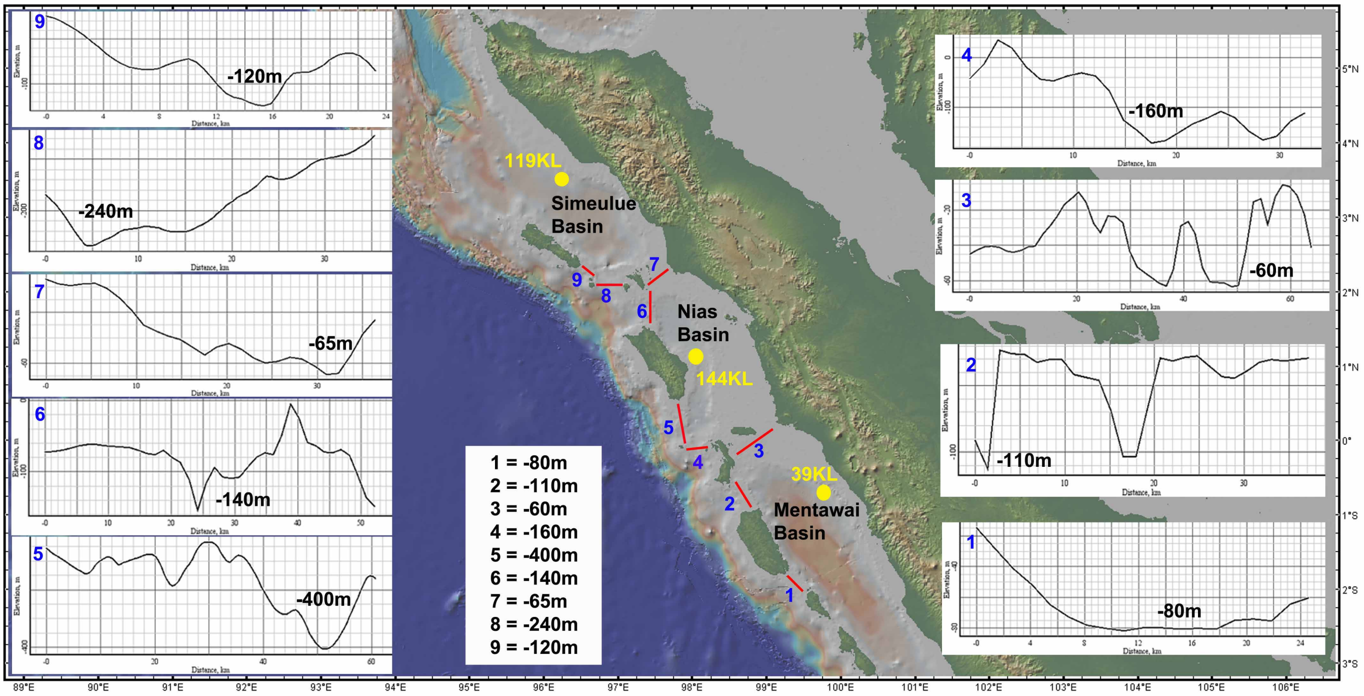
55. Wang, L. Isotopic signals in two morphotypes of *Globigerinoides ruber* (white) from the South China Sea: implications for monsoon climate change during the last glacial cycle. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **161,** 381–394 (2000).
56. Mohtadi, M. *et al.* Low-latitude control on seasonal and interannual changes in planktonic foraminiferal flux and shell geochemistry off south Java: a sediment trap study. *Paleoceanography* **24,** PA1201 (2009).
57. Barker, S., Greaves, M. & Elderfield, H. A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry. *Geochem. Geophys. Geosyst.* **4,** 8407 (2003).
58. Greaves, M. *et al.* Interlaboratory comparison study of calibration standards for foraminiferal Mg/Ca thermometry. *Geochem. Geophys. Geosyst.* **9,** Q08010 (2008).
59. Anand, P., Elderfield, H. & Conte, M. H. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18,** 1050 (2003).
60. Bemis, B. E., Spero, H. J., Bijma, J. & Lea, D. W. Reevaluation of the oxygen isotopic composition of planktonic foraminifera: experimental results and revised paleotemperature equations. *Paleoceanography* **13,** 150–160 (1998).
61. Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21,** 295–305 (2002).
62. Bevington, P. R. & Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences* 3rd edn, 320 (McGraw-Hill, 2003).
63. Mudelsee, M. Ramp function regression: a tool for quantifying climate transitions. *Comput. Geosci.* **26,** 293–307 (2000).
64. Fleitmann, D. *et al.* Holocene forcing of the Indian monsoon recorded in a stalagmite from Southern Oman. *Science* **300,** 1737–1739 (2003).
65. Fleitmann, D. *et al.* Timing and climatic impact of Greenland interstadials recorded in stalagmites from northern Turkey. *Geophys. Res. Lett.* **36,** L19707 (2009).
66. Steffensen, J. P. *et al.* High-resolution Greenland ice core data show abrupt climate change happens in few years. *Science* **321,** 680–684 (2008).
67. Mudelsee, M. & Raymo, M. E. Slow dynamics of the Northern Hemisphere glaciation. *Paleoceanography* **20,** PA4022 (2005).
68. Mudelsee, M. Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Math. Geol.* **35,** 651–665 (2003).
69. Collins, W. D. *et al.* The Community Climate System Model version 3 (CCSM3). *J. Clim.* **19,** 2122–2143 (2006).
70. Yeager, S. G., Shields, C. A., Large, W. G. & Hack, J. J. The low-resolution CCSM3. *J. Clim.* **19,** 2545–2566 (2006).
71. Levis, S., Bonan, G. B., Vertenstein, M. & Oleson, K. W. *The Community Land Model's Dynamic Global Vegetation Model (CLM-DGVM): Technical Description and User's Guide* (National Center for Atmospheric Research, 2004).
72. Oleson, K. *et al. Technical Description of the Community Land Model (CLM).* (National Center for Atmospheric Research, 2004).
73. Oleson, K. W. *et al.* Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res.* **113,** G01021 (2008).
74. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features. *Clim. Past* **3,** 261–277 (2007).
75. Berger, A. L. Long-term variations of daily insolation and Quaternary climate changes. *J. Atmos. Sci.* **35,** 2362–2367 (1978).
76. Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32,** 111–149 (2004).
77. Merkel, U., Prange, M. & Schulz, M. ENSO variability and teleconnections during glacial climates. *Quat. Sci. Rev.* **29,** 86–100 (2010).
78. Flückiger, J. *et al.* N_2O and CH_4 variations during the last glacial epoch: insight into global processes. *Glob. Biogeochem. Cycles* **18,** GB1020 (2004).
79. Spahni, R. *et al.* Atmospheric methane and nitrous oxide of the Late Pleistocene from Antarctic ice cores. *Science* **310,** 1317–1321 (2005).
80. Ahn, J. & Brook, E. J. Atmospheric CO_2 and climate from 65 to 30 ka B.P. *Geophys. Res. Lett.* **34,** L10703 (2007).
81. Otto-Bliesner, B. L. *et al.* Climate sensitivity of moderate- and low-resolution versions of CCSM3 to preindustrial forcings. *J. Clim.* **19,** 2567–2583 (2006).
82. Ding, Q. & Wang, B. Circumglobal teleconnection in the Northern Hemisphere summer. *J. Clim.* **18,** 3483–3505 (2005).
83. Ding, Q. & Wang, B. Intraseasonal teleconnection between the summer Eurasian wave train and the Indian Monsoon. *J. Clim.* **20,** 3751–3767 (2007).
84. Broccoli, A. J., Dahl, K. A. & Stouffer, R. J. Response of the ITCZ to Northern Hemisphere cooling. *Geophys. Res. Lett.* **33,** L01702 (2006).



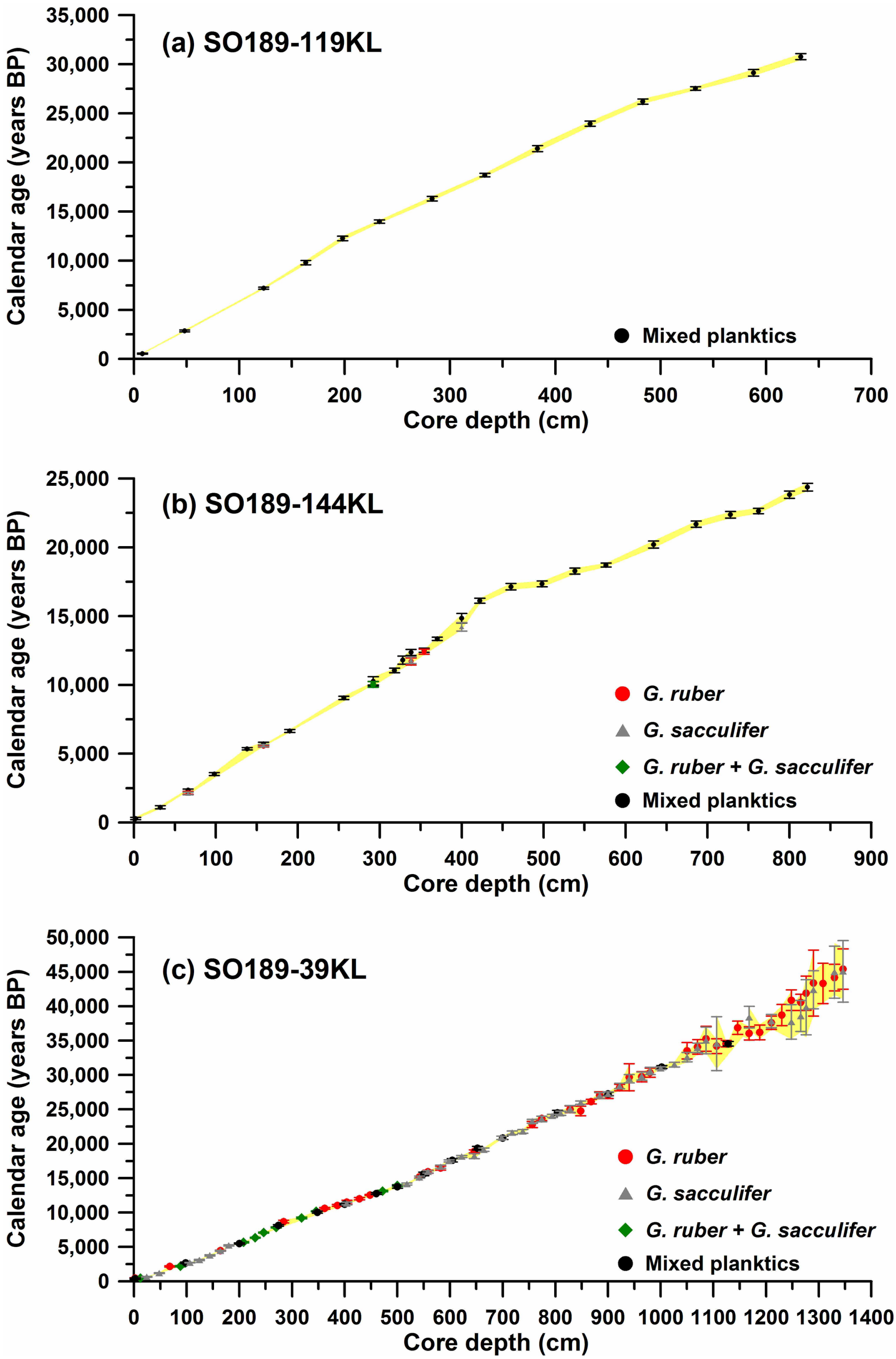
Extended Data Figure 1 | Instrumental records of temperature and precipitation in the study area. Records are at or close to the sites 119KL (black) and 39KL (red). **a**, Average monthly SST for the Simeulue basin (4° N, 96° E; black) and the northern Mentawai basin (2° S, 100° E; red) based on extended reconstruction sea surface temperature (ERSST) data from 1854 to 2008 (http://nomads.ncdc.noaa.gov/las/getUI.do). Dashed lines indicate average SST for the entire period. **b**, Twenty-four-hour air temperatures measured in Sabang in northwestern Sumatra (from 1976 to 1989; black) and in Padang in western Sumatra (from 1850 to 1989; red; http://climexp.knmi.nl). Dashed lines indicate the average air temperature over the entire period. **c**, Average monthly precipitation (mm per month) over Banda Aceh in northwestern Sumatra (black) and Padang (red), between 1879 and 1989 (http://climexp.knmi.nl). Open circles represent mean monthly precipitation of different seasons (winter, spring, summer and autumn), with the numbers indicating the percentage contribution of each season to the total annual precipitation. Dashed lines indicate average monthly precipitation for the entire period. Note the small seasonality of SST, air temperature and precipitation in the study area.



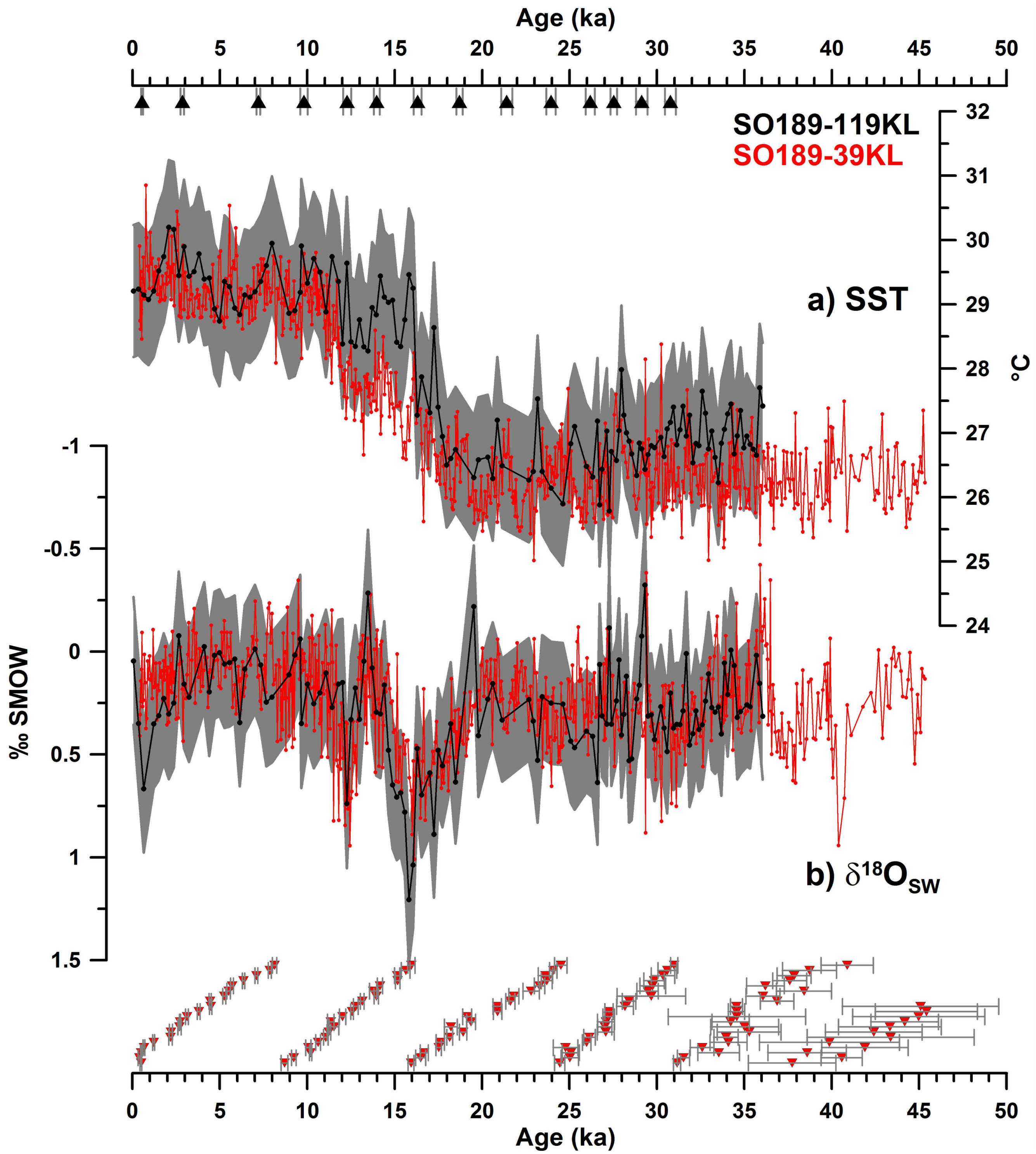
Extended Data Figure 2 | Seasonality of surface currents, SST and salinity in the eastern Indian Ocean. Seasonal changes in SST (colour shading), salinity (dashed lines; p.s.u.) and surface currents (arrows) in the study area during boreal summer (top) and winter (bottom). The meridional Ekman transport (ME) is also indicated with arrows. Seasonal SST is averaged for the period between 2002 and 2010 (http://oceancolor.gsfc.nasa.gov/cgi/l3). Salinities are averaged for the period between 1960 and 2004[31]. Surface currents and ME are redrawn following ref. 35. Note the seasonal reversal of the surface currents and the MET, and the small seasonality of SST and salinity off western and northwestern Sumatra. The positions of the cores from the tropical eastern Indian Ocean are indicated by stars (this study). NECC, north equatorial counter current; NMC, northeast monsoon current; SECC, south equatorial counter current; SJC, south Java current; SMC, southwest monsoon current.



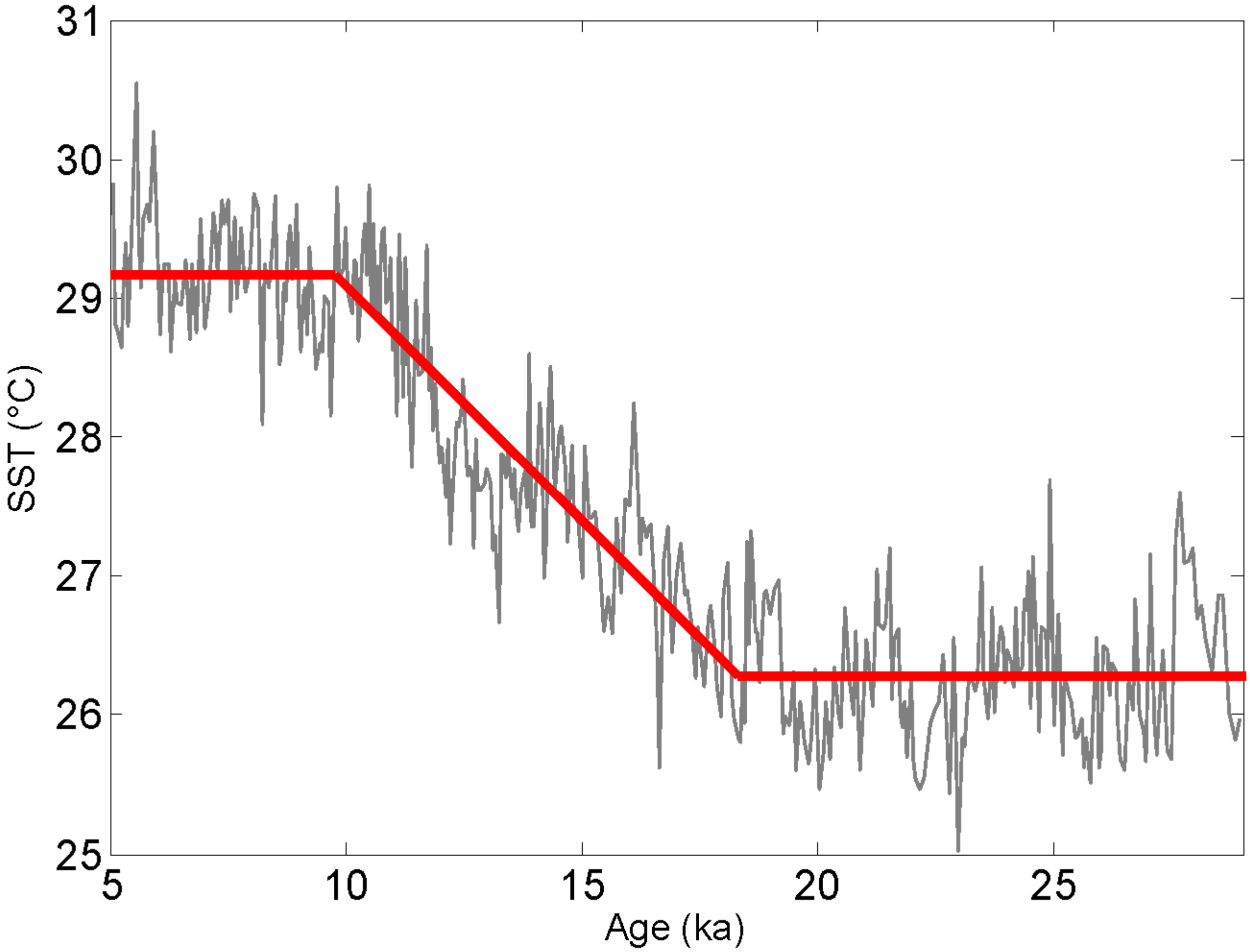
Extended Data Figure 3 | Sill depths in the study area. Sill depths of the Simeulue basin (1–3), the Nias basin (4–6) and the northern Mentawai basin (7–9), with the positions of the cores indicated (yellow dots). The maximum depth of each sill is as indicated.



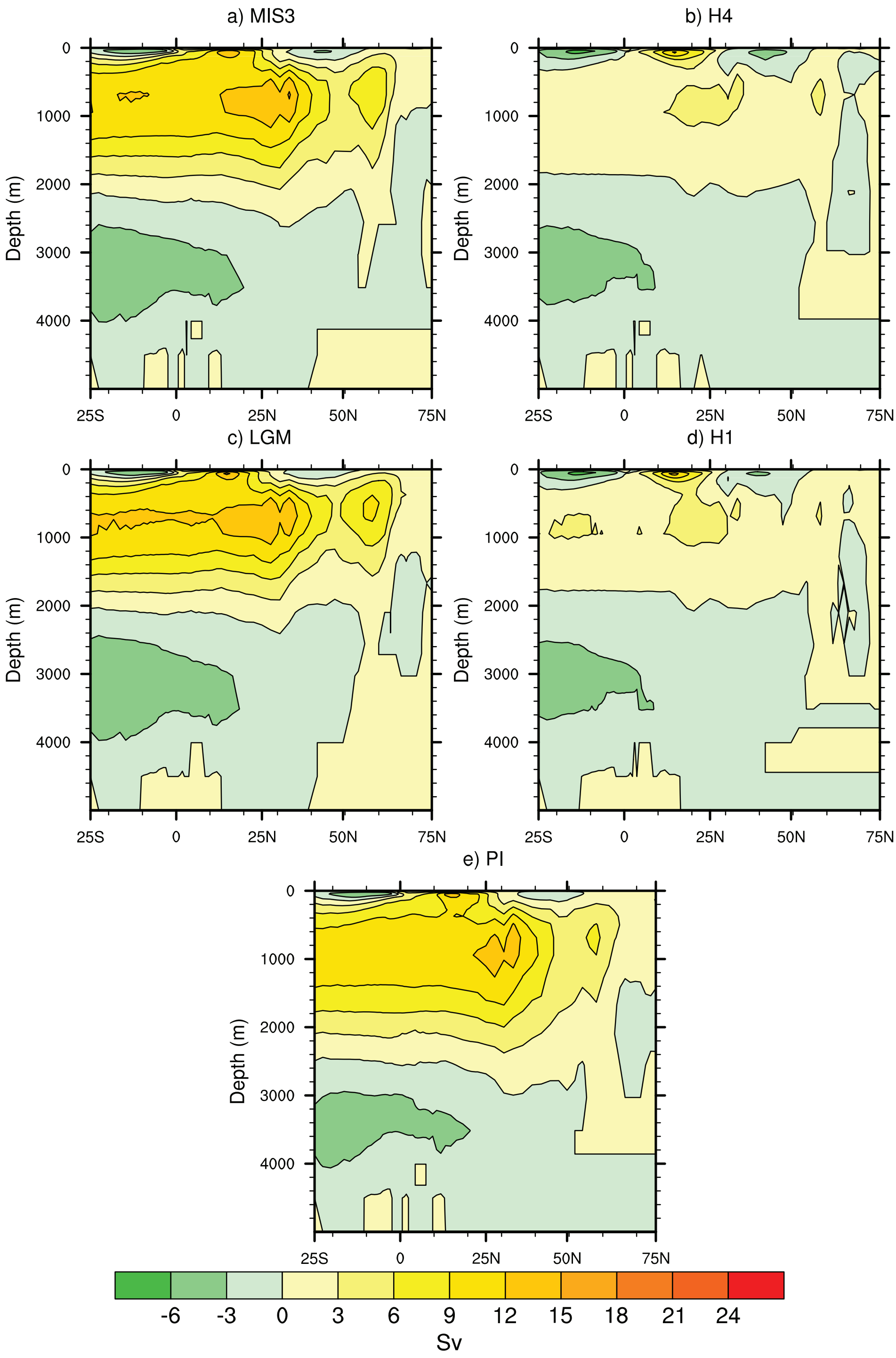
Extended Data Figure 4 | Age–depth relationship of the investigated cores. Core depth (cm) versus calendar age (years) with 2σ errors (bars and yellow envelope) in cores 119KL (**a**), 144KL (**b**) and 39KL (**c**).



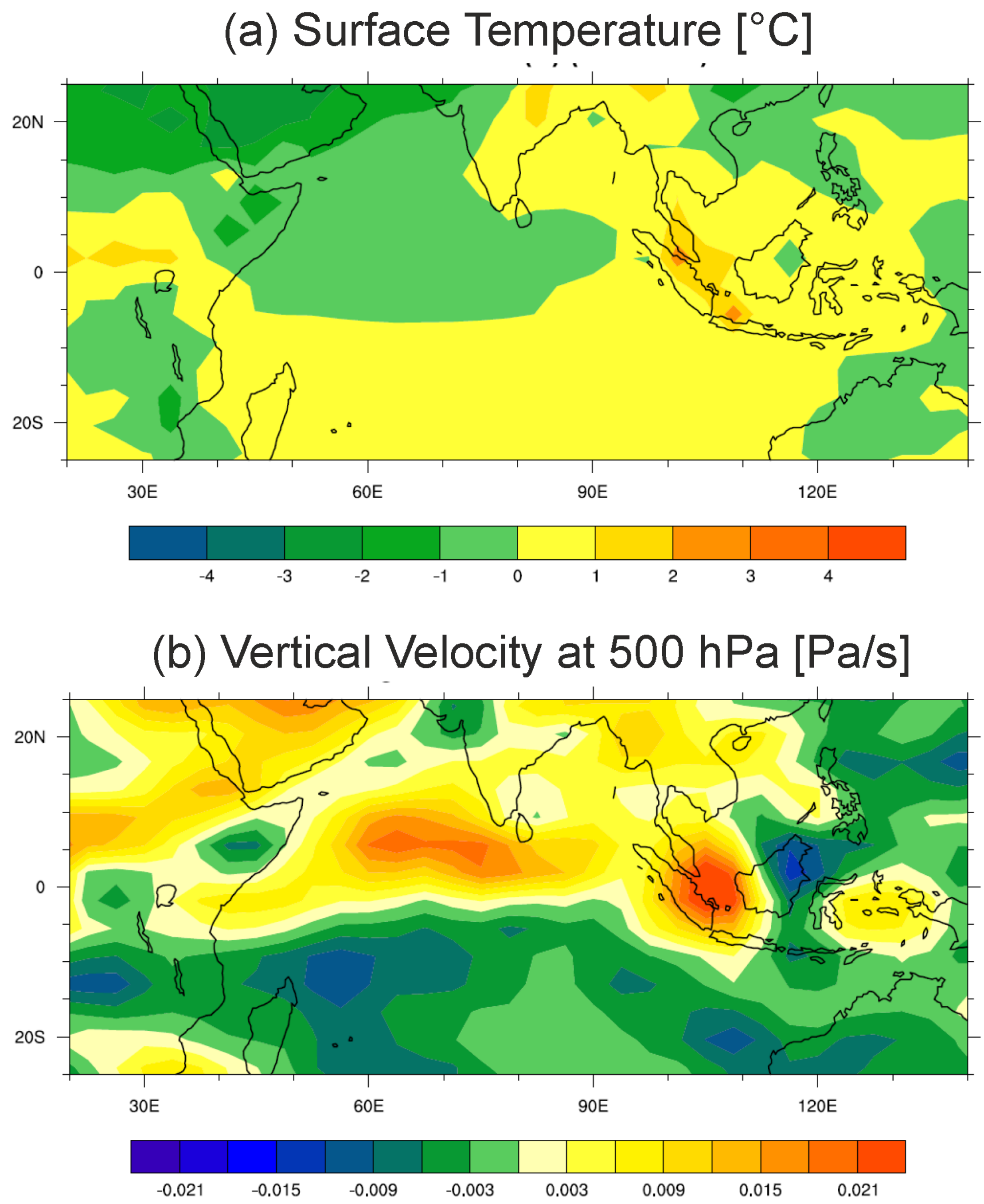
Extended Data Figure 5 | Estimated errors (1σ) for SST and $\delta^{18}O_{SW}$ in core 119KL. Grey envelopes indicate errors in reconstructions of SST (**a**) and $\delta^{18}O_{SW}$ (**b**). For comparison, the 39KL records (red) are shown. Grey bars indicate 2σ errors of the calibrated radiocarbon ages (black and red triangles).



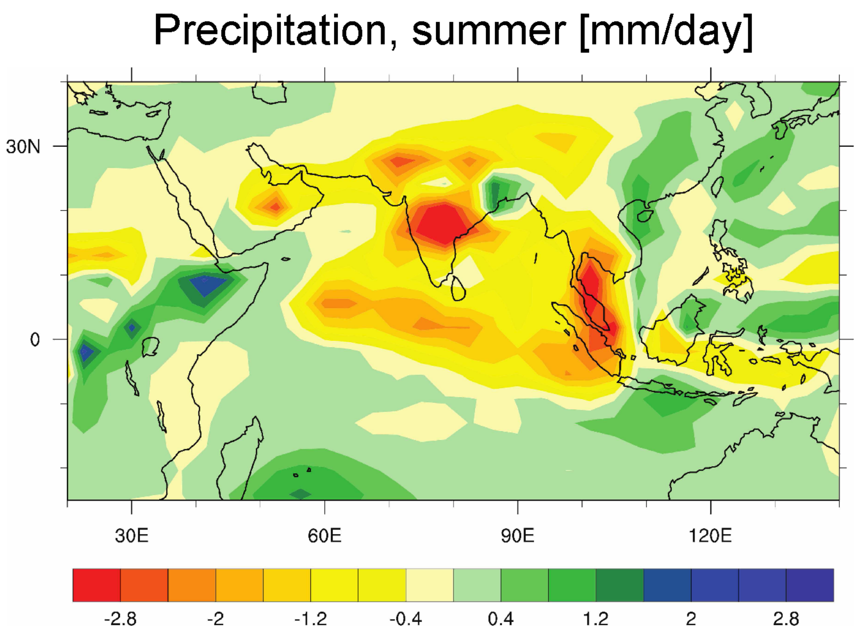
Extended Data Figure 6 | Mg/Ca SST record of core 39KL for the period 5–29 kyr ago, along with the fitted ramp function (red).



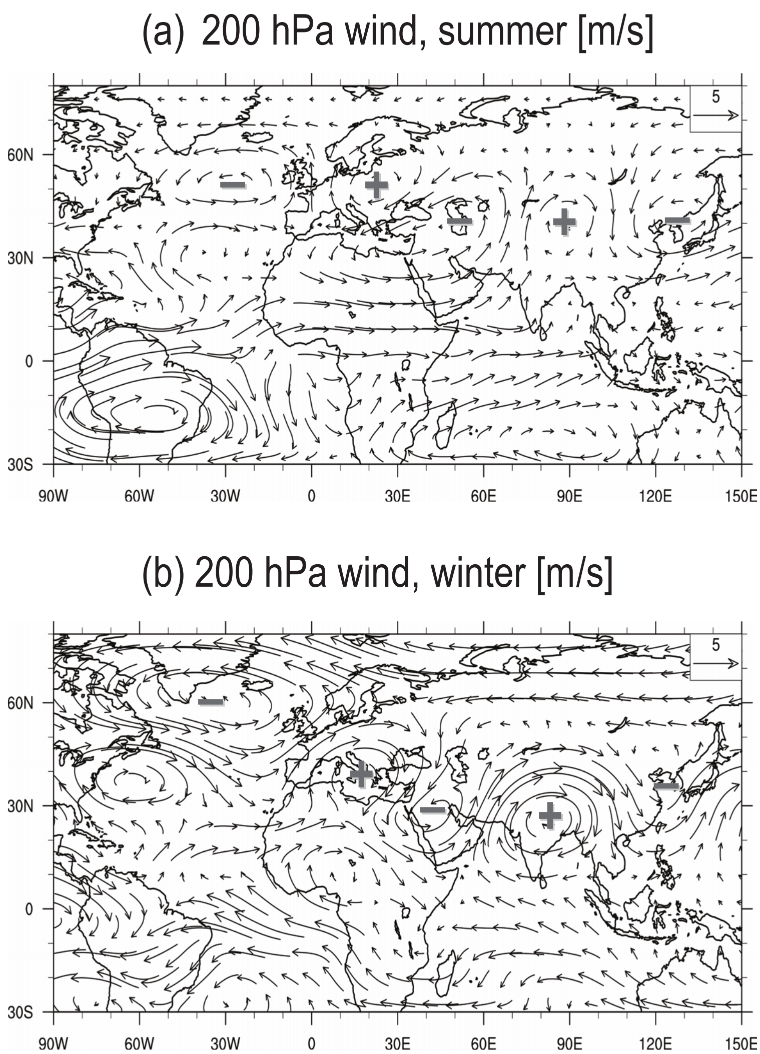
Extended Data Figure 7 | AMOC for different climate states, as simulated by CCSM3. Meridional overturning stream function averaged over the last 100 yr of each experiment for the MIS3 baseline run (**a**), the H4 hosing experiment (**b**), the LGM simulation (**c**), the H1 hosing experiment (**d**) and the pre-industrial control run (**e**).



Extended Data Figure 8 | Climatic response to a substantial slowdown of the AMOC under LGM (21 kyr ago) boundary conditions in a CCSM3 simulation. Shown are long-term (100-yr) annual means of climatic anomalies (Heinrich stadial 1 hosing experiments minus LGM baseline run) for surface temperature (**a**) and vertical velocity (**b**) at 500 hPa.



Extended Data Figure 9 | Summer (June, July and August) precipitation response to a substantial slowdown of the AMOC under MIS3 (38 kyr ago) boundary conditions, as simulated by CCSM3. Shown are 100-yr averages (Heinrich stadial 4 hosing experiment minus MIS3 baseline run).



Extended Data Figure 10 | Upper-tropospheric (200 hPa) wind response to a substantial slowdown of the AMOC (Heinrich stadial 4 hosing experiment minus MIS3 baseline run; 100-yr averages). **a**, Summer (June, July and August) response; **b**, winter (December, January and February) response. Wave trains are highlighted by plus symbols (positive geopotential height anomaly/anticyclonic circulation anomaly) and minus symbols (negative geopotential height anomaly/cyclonic circulation anomaly).

LETTER

doi:10.1038/nature13245

Electrical conductivity during incipient melting in the oceanic low-velocity zone

David Sifré[1,2,3], Emmanuel Gardés[1,2,3,4], Malcolm Massuyeau[1,2,3], Leila Hashim[1,2,3], Saswata Hier-Majumder[1,2,3,5,6] & Fabrice Gaillard[1,2,3]

The low-viscosity layer in the upper mantle, the asthenosphere, is a requirement for plate tectonics[1]. The seismic low velocities and the high electrical conductivities of the asthenosphere are attributed either to subsolidus, water-related defects in olivine minerals[2–4] or to a few volume per cent of partial melt[5–8], but these two interpretations have two shortcomings. First, the amount of water stored in olivine is not expected to be higher than 50 parts per million owing to partitioning with other mantle phases[9] (including pargasite amphibole at moderate temperatures[10]) and partial melting at high temperatures[9]. Second, elevated melt volume fractions are impeded by the temperatures prevailing in the asthenosphere, which are too low, and by the melt mobility, which is high and can lead to gravitational segregation[11,12]. Here we determine the electrical conductivity of carbon-dioxide-rich and water-rich melts, typically produced at the onset of mantle melting. Electrical conductivity increases modestly with moderate amounts of water and carbon dioxide, but it increases drastically once the carbon dioxide content exceeds six weight per cent in the melt. Incipient melts, long-expected to prevail in the asthenosphere[10,13–15], can therefore produce high electrical conductivities there. Taking into account variable degrees of depletion of the mantle in water and carbon dioxide, and their effect on the petrology of incipient melting, we calculated conductivity profiles across the asthenosphere for various tectonic plate ages. Several electrical discontinuities are predicted and match geophysical observations in a consistent petrological and geochemical framework. In moderately aged plates (more than five million years old), incipient melts probably trigger both the seismic low velocities and the high electrical conductivities in the upper part of the asthenosphere, whereas in young plates[4], where seamount volcanism occurs[6], a higher degree of melting is expected.

The lithosphere is a chemically depleted and mechanically strong region of the uppermost mantle, overlying the chemically enriched and mechanically weak asthenosphere[1–3,8,16]. Volatile enrichments in the asthenosphere have long been known to trigger incipient melting[10,13–15] (that is, a small degree of partial melting due to small amounts of CO_2 and H_2O) in the upper part of the asthenosphere, and a link between incipient melting and the seismic low-velocity zone has also long been suggested[13,15,17,18]. Here we demonstrate that incipient melting of the mantle can also trigger high electrical conductivities. We assume that the low-viscosity layer, the layer of high electrical conductivity and the layer of low seismic velocity are coincident, and use the term 'asthenosphere' for this layer.

Although the characteristics of the asthenosphere are commonly linked to H_2O-related defects in olivine[2–4], a number of multidisciplinary observations[5–8,10,13,15,17–20] and the discovery of petit-spot volcanoes[21] (small volcanic edifices found on old oceanic lithosphere) indicate that the asthenosphere most probably contains partial melt.

Two issues arise when the observed features of the Asthenosphere are attributed to partial melting. First, a few volume per cent of basaltic melt are generally required to explain high electrical conductivities, which is problematic because melt would unavoidably tend to rise if present at such high amounts[11,12]. Second, the lithosphere–asthenosphere boundary (LAB) occurs at a near-constant depth of 50–75 km for both warm, young lithospheres and cold, ancient lithospheres[4,5,8,16,18], and in most cases the temperature may not be sufficiently high to produce such amounts of melt.

However, incipient CO_2- and H_2O-rich melts, which are stable under the pressure (P)–temperature (T)–fugacity (f_{O_2}) conditions of the asthenosphere[10,13–15,17,22,23], allow melting in both warm and cold regions of the asthenosphere[17]. Low-temperature carbonatite melts, composed of almost 50% CO_2, are characterized by high electrical conductivities[19], but their stability is restricted to the coldest and driest regions of the asthenosphere[17]. Increasing temperature or H_2O content changes the composition of the prevailing melts to intermediates between basalts and carbonatites, often described as carbonated basalts[17,22]. Very little is known about the physical properties of such intermediate volatile (CO_2 and H_2O)-rich melts. In particular, their electrical properties have never

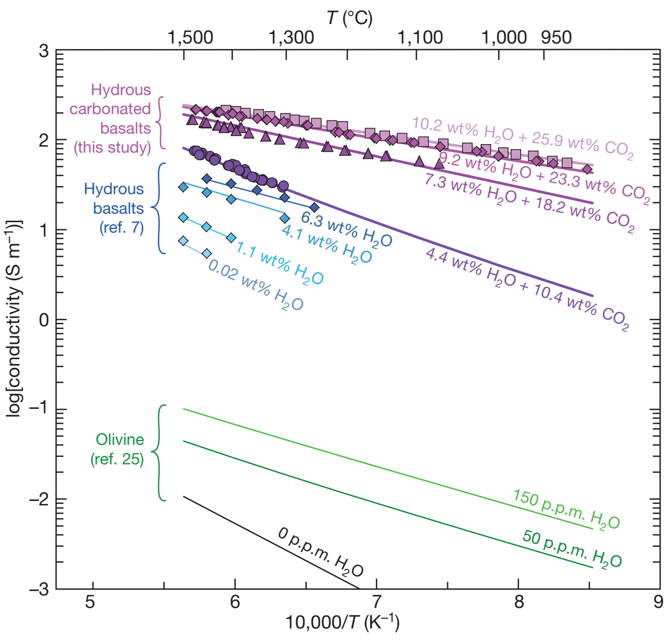


Figure 1 | Electrical conductivity of hydrous carbonated basalts, hydrated basalts and hydrous olivine. The conductivities of the hydrous carbonated basalts experimentally measured in this study are by far the highest, reaching up to 200 S m^{-1} and being about one and four order of magnitude higher than hydrated basalts[7] and hydrous olivine[25], respectively. The fitting curves are calculated according to our conductivity model for CO_2- and H_2O-bearing melts (equation (1)).

[1]Université d'Orléans, Institut des Sciences de la Terre d'Orléans (ISTO), Unité mixte de recherche (UMR) 7327, 45071 Orléans, France. [2]Centre National de la Recherche Scientifique (CNRS)/ISTO, UMR 7327, 45071 Orléans, France. [3]Bureau des Recherches Géologiques et Minières, ISTO, UMR 7327, BP 36009, 45060 Orléans, France. [4]Commissariat à l'Energie Atomique-CNRS-École nationale supérieure d'ingénieurs de Caen-Université de Caen Basse Normandie, Centre de Recherche sur les Ions, les Matériaux et la Photonique, UMR 6252, BP 5133, 14070 Caen, France. [5]Department of Geology and Center for Scientific Computation and Applied Mathematical Modeling, University of Maryland, Maryland 20742, USA. [6]Department of Earth Sciences, Royal Holloway University of London, Egham TW20 0EX, UK.

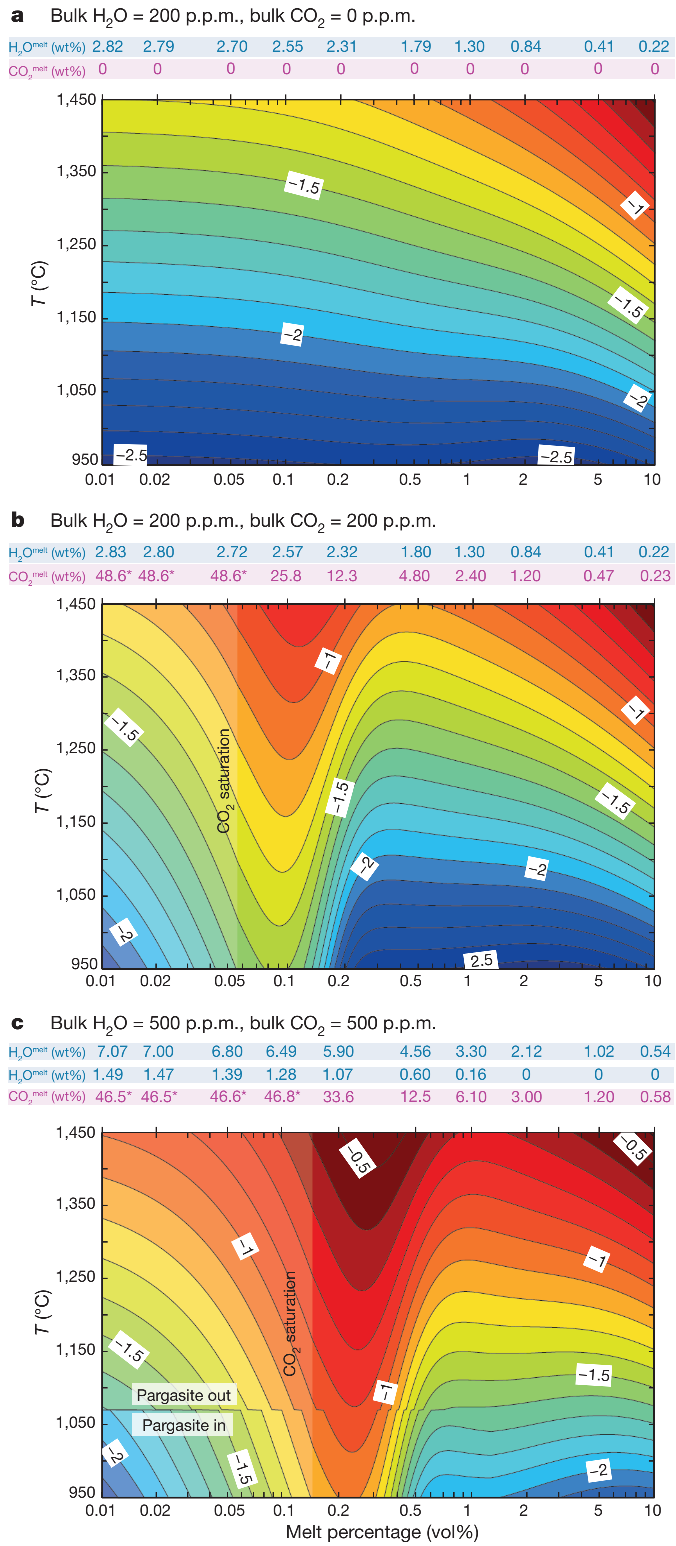


Figure 2 | Effect of incipient melt on the electrical conductivity of depleted and enriched carbonated peridotites. The conductivity of partially molten peridotite (log values increasing from cold to warm colours) is reported as a function of melt content and temperature for CO_2-free peridotite with 200 p.p.m. H_2O (**a**) and depleted and enriched CO_2-bearing hydrous systems (**b**, **c**). H_2O partitions between minerals and melt, and CO_2 distributes in melt only (Methods). The H_2O and CO_2 contents in the melt are shown at the top of each figure. Addition of CO_2 triggers a peak in conductivity at 0.1–0.3 vol% of melt, where the intergranular liquid is CO_2 rich and therefore highly conductive. At higher degrees of melting, the bulk conductivity decreases because volatiles are diluted in the melt (melt water and CO_2 are tabulated above each panel), which becomes basaltic. A peridotite with 0.1 vol% carbonated basalt is as conductive as one with 10 vol% basalt. The two sets of melt H_2O contents given in **c** correspond to pargasite-saturated ($T < 1{,}070$ °C, bottom line) and pargasite-undersaturated (1,070°C, top line) melt.

been measured. To address the issues regarding onset of partial melting at the LAB, and to permit a test of the incipient melting model suggested by petrological studies[15,17], we performed electrical conductivity measurements on CO_2- and H_2O-rich melts.

We developed an experimental set-up specifically adapted for liquids with high conductivities (Extended Data Fig. 1). A high-performance four-wire method was used under high-pressure and high-temperature conditions in a piston–cylinder apparatus (Methods and Extended Data Figs 1, 2 and 3a). Five melts, with CO_2 and H_2O contents ranging from 10 to 48 wt% and 0 to 10 wt%, respectively, were analysed by impedance spectroscopy in the temperature range 900–1,500 °C at a confining pressure of 3 GPa. We tested the reproducibility of the measurements by taking measurements during both cooling and heating of the samples (Extended Data Fig. 3a), and we verified that decarbonation and dehydration of samples at high temperature did not affect the conductivity results. Figure 1 reports the measured electrical conductivities as a function of reciprocal temperature. For similar H_2O contents, the electrical conductivity of carbonated basalts is higher than that of hydrated basalts, and the difference increases with an increase in the CO_2 content of the melt to a maximum of nearly one log unit (a factor to ten). The most CO_2-rich melt has conductivities higher than 200 S m^{-1}. We develop a semi-empirical model that takes into account the two parallel conductive processes operating in carbonated basalts: conduction by covalent, polymer-like hydrous silicate melts and ionic conduction by carbonate melts[19]. The corresponding conductivity is given by

$$\sigma_{\mathrm{model}} = \sigma^{H_2O} + \sigma^{CO_2}$$
$$= \sigma_0^{H_2O} \exp\left(\frac{-E_a^{H_2O}}{RT}\right) + \sigma_0^{CO_2} \exp\left(\frac{-E_a^{CO_2}}{RT}\right) \qquad (1)$$

Here T is the temperature, R is the gas constant, and σ_0 and E_a respectively stand for the pre-exponential and the activation energy terms of the two Arrhenius laws defining the electrical contributions of the hydrous silicate (σ^{H_2O}) and the carbonate (σ^{CO_2}). The conductivities calculated using equation (1), as shown in Fig. 1, reproduce our measurements and those of ref. 7 on CO_2-free hydrated basalts with an average precision of 5% (Methods). The effect of CO_2 on melt conductivity predicted by equation (1) is negligible at low CO_2 content, but increases sharply for CO_2 content higher than 6 wt%. Such a change is most probably caused by an abrupt transition in the melt structure and properties from silicate type to carbonate type.

We calculate the mantle electrical conductivity for variable amounts of bulk H_2O and CO_2 contents in a partially molten peridotite. We assume that the interconnected melt is equally distributed between grain-edge tubules and grain-boundary melt films[18,20,24] (Methods). The conductivity of hydrated olivine was calculated from ref. 25, and equation (1) was used for CO_2- and H_2O-bearing melts. We assume that carbon is exclusively soluble in the melt[23] (carbonate units) and computed the partitioning of H_2O between carbonated melt, pargasite, olivine and peridotite, combining refs 9, 10 and 22. We report results for partially molten peridotite containing only H_2O (Fig. 2a) and both CO_2 and H_2O (Fig. 2b, c). In all simulations, partitioning constraints for CO_2 and H_2O between solids and melts require that CO_2- and H_2O-rich melts can be produced only at the onset of mantle partial melting and that small melt fractions always contain much more CO_2 than H_2O (Fig. 2, top axis). If melting of more than 1% is attained, the melt volatile contents drop to values that modestly affect their electrical conductivity.

The CO_2-free depleted mantle, containing about 200 p.p.m. H_2O (refs 26–28), cannot be conductive at temperatures below 1,350 °C (that is,

$\sigma \geq 0.1\ S\ m^{-1}$) unless it contains more than ~5 vol% basaltic melt (Fig. 2a). Only unreasonably high temperatures for the LAB (>1,450 °C) can make the mantle conductive with small amounts of melt (<1 vol%). Moreover, at high melt percentage, H_2O has almost no effect on mantle conductivity, because its content in the melt remains small (that is, less than 1 wt% H_2O negligibly affects basalt conductivity). If an H_2O-enriched (500 p.p.m.) mantle is considered then a reasonably low melt content (1 vol%) can cause high conductivity, but it still requires a high temperature (>1,325 °C; Extended Data Fig. 4) and the enriched mantle is also CO_2 rich[26–28].

In the presence of CO_2, the formation of incipient CO_2-rich melts (<0.5 vol%) disproportionately increases the effective electrical conductivity of the mantle (Fig. 2b, c). For example, in the depleted mantle, containing 200 p.p.m. H_2O and 200 p.p.m. CO_2 (ref. 26) and fuelling the dominant part of mid-ocean-ridge basalts[27,28], 0.1–0.15 vol% of melt at 1,325 °C can explain the existence of a layer of high electrical conductivity reported in oceanic domains[5,18,29]. The melt is a carbonated basalt, typically containing 15–35 wt% CO_2 and about 2–3 wt% H_2O (Fig. 2b). Remarkably, the enriched mantle[27,28] with 500 p.p.m. H_2O and 500 p.p.m. CO_2 can produce high conductivities at temperature and melt fractions as low as 1,050 °C and 0.2 vol%, respectively. We also notice that incipient melting of the enriched mantle triggers conductivities that are 2.5 times greater than those in the depleted mantle, making variations in electrical conductivity a powerful probe of the chemical enrichment in the upper mantle.

The stability of incipient melts in the upper part of the asthenosphere is expected as a result of petrological constraints[13–15] (not considered in Fig. 2). The *P*–*T* region of incipient melting in peridotite is shown in Extended Data Fig. 5 together with the stability domain of pargasite, which is the main solid host for H_2O in peridotite containing more than 150–200 p.p.m. H_2O (ref. 10). We calculate that the presence of pargasite restricts the amount of H_2O to 40–50 p.p.m. in olivine, according to partition coefficients among peridotite minerals[9,10]. Pargasite is, however, stable at temperatures less than 1,070 °C (ref. 10) and for the enriched mantle (>200 p.p.m. H_2O). On the basis of the *P*–*T* phase diagram of Extended Data Fig. 5, and considering oceanic geotherms at 23.5, 35 and 70 Myr ago, we have calculated one-dimensional conductivity profiles illustrating the impact of several petrological discontinuities (Fig. 3). We have considered the depleted mantle (200 p.p.m. H_2O and 100–500 p.p.m. CO_2) and the enriched mantle (500 p.p.m. H_2O and 500 p.p.m. CO_2). Variable CO_2 contents in the depleted mantle account for the fact that mid-ocean-ridge basalts have degassed their CO_2 with the result that the carbon contents of their sources are highly uncertain[26–28].

The upper discontinuity (Fig. 3) predicted by our model is the beginning of incipient melting at ~50-km depth for young, warm plates and at ~70 km for colder, older plates. This discontinuity marks the thermodynamic boundary between CO_2-rich melt and CO_2-rich vapour[14], the melt being stable at greater depth. In the case of an enriched mantle, an additional discontinuity occurs owing to the pargasite dehydration melting reaction (producing CO_2- and H_2O-rich, low-SiO_2 melt). This discontinuity can be shallower than the previously described discontinuity for young (23.5-Myr-old) plates and deeper for old (70-Myr-old) plates. For 35-Myr-old plates, the two discontinuities occur at the same depth (60 km). The lowest discontinuity shown in Fig. 3 occurs in the depth interval 120–150 km and is described as the region of redox melting[22,23]; this is the boundary separating diamonds from CO_2-rich melts, the melt being stable at shallower depth. Incipient melting, which triggers the conductive region of the asthenosphere, is therefore permitted between the redox-melting lower boundary and the decarbonation upper boundary, and this agrees well with electromagnetic observations in oceanic domains[4,5,18,22,29], although ref. 29 indicates slightly deeper ranges.

The increase in conductivity in the incipient melting region is major, being half a log unit for the depleted mantle (200 p.p.m. CO_2) and more than one log unit for the mantle containing 500 p.p.m. CO_2. High conductivities of $0.1\ S\ m^{-1}$ or more can be reached for CO_2 contents as low as 300 p.p.m. in the case of young plates. We note the surprising effect of H_2O (Fig. 3): incipient melting in a mantle with 500 p.p.m. H_2O and 500 p.p.m. CO_2 induces lower conductivities than in a mantle with

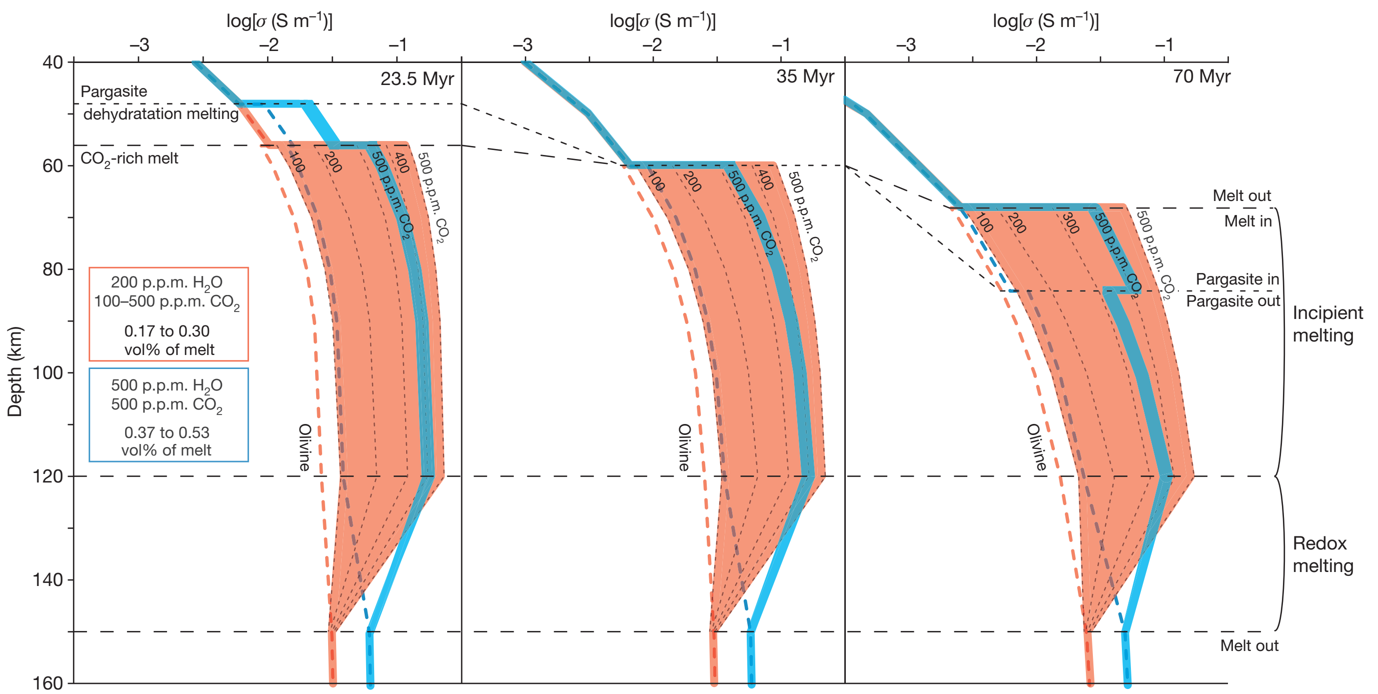


Figure 3 | Petrologically based conductivity profiles across the incipient melting region under the LAB for various ages. The top axis indicates electrical conductivity and how it varies with depth during cooling of the lithosphere for ages of 23.5, 35 and 70 Myr (see choices of geotherm in Fig. 4). Conductivities were calculated using the same model as in Fig. 2 (Methods). Several electrical discontinuities are predicted at variable depths, on the basis of the phase-equilibria relationships shown in Extended Data Figs 5 and 6; the most striking conductivity jump is related to the upper and lower boundaries of incipient melting (55–150 km). The volatile-depleted and -enriched mantles are considered, and it can be seen that the conductivity during incipient melting is strongly correlated to CO_2 contents (grey dashed lines labelled from 100 to 500 p.p.m.).

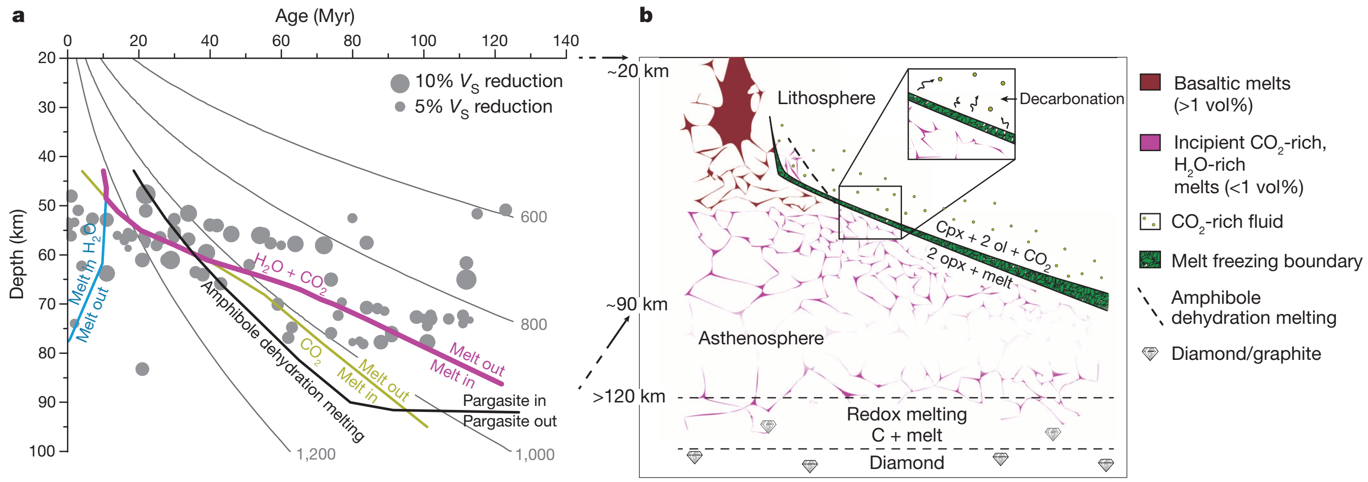


Figure 4 | The oceanic seismic low-velocity zone bracketed by the upper and lower boundaries of incipient melting. **a**, Oceanic crustal ages versus depth of seismic discontinuities (S-wave (Vs) reductions) marking the LAB (grey circles[8]) beneath the Pacific Ocean. Colour curves designate the solidi for hydrated (200 p.p.m. H_2O; blue), carbonated (green) and H_2O-undersaturated carbonated (purple) peridotites. Isotherms (grey curves) are calculated from a model of sudden half-space cooling, assuming[8] $\Delta T = 1{,}350\ ^{\circ}C$, an average plate velocity of 8 cm yr^{-1} and a thermal diffusivity of 1 $mm^2\ s^{-1}$. Varying the plate velocity does not change the plot. **b**, A visual picture capturing the domain of incipient melting in the oceanic low-velocity zone. This zone is bounded from below by the redox melting[22,23] and from above by the decarbonation[14] leading to the freezing of incipient melts. This boundary constitutes an impermeable layer leaving a clinopyroxene-rich residue and a CO_2-rich vapour phase.

200 p.p.m. H_2O and 500 p.p.m. CO_2. The imaging of the 23.5-Myr-old LAB by at ~50-km depth revealed conductivities of 0.1–0.2 S m^{-1} (ref. 5). These are definitely not explicable by melting of a CO_2-free, H_2O-depleted mantle, because the required temperatures or melt contents, or both, are too great (Fig. 2). They cannot be explained by pargasite dehydration melting in a CO_2-free, H_2O-enriched mantle either, because this process cannot produce high enough conductivities (Fig. 3 and Extended Data Fig. 4). Once deciphered in a petrological framework, the conductivities of the LAB in ref. 5 can be reached by incipient melting of a mantle containing 400 p.p.m. CO_2 (Fig. 3). We note that ref. 5 introduced moderate electrical anisotropy in the inversion of their magnetotelluric data, whereas we merely discuss here the geometric mean conductivity, which is much less model dependent.

The presence of CO_2-rich melts in the asthenosphere not only better explains the electrical properties of the asthenosphere, but also explains the weak dependence of the lithosphere thickness on the age of the oceanic crust (Fig. 4). The bottom of the lithosphere in Fig. 4 marks a seismic discontinuity characterized by a reduction in S-wave velocity of 5–15% (refs 8, 16). This discontinuity cannot be caused by partial melting of adry or H_2O-undersaturated mantle, because such conditions are possible only at greater depths and temperatures[9] (Fig. 4, blue melting curve). Previously suggested melting reactions such as the dehydration melting of amphibole[10] also fail to reproduce the depth–age relationships of the LAB (Fig. 4). Remarkably, the CO_2–H_2O melting curve[14], which delimits the upper boundary of the incipient melting region already shown in Fig. 3, ranges in depth from 50 to 80 km from the youngest to the oldest lithospheres (Fig. 4, purple curve). This correlates fairly well with the bottom of the lithosphere as imaged by the seismic discontinuity. The lower limit of incipient melting, that is, the redox melting[22,23], also well matches the lower part of the seismic low-velocity zone[8] at depths of about 140–180 km. At low pressures, above the incipient melting region, the decarbonation of the melt forms an impermeable layer in which buoyant CO_2-rich melts are frozen into clinopyroxene-rich residue (with pargasite) and CO_2-rich fluids (Fig. 4). Melting is therefore permitted in the asthenosphere and the melt cannot rise through the LAB because of the existence of this melt-freezing boundary (Methods). It is only where the mantle is hot enough to suppress the freezing reaction (such that melting no longer requires CO_2) and where melt fractions are large enough[12] (2–5 vol%) that melts can rise through the LAB. This occurs for young plates (<5 Myr old), where volcanic seamounts are observed, and has been related[6] to the electrical properties of the young LAB[4].

Incipient melting has long been described as a key petrological process operating in the seismic low-velocity region marking the upper asthenosphere[13,15,17]. The mantle geochemistry and petrology in this region argue for production of incipient CO_2-enriched melts[9,10,13,15]. We demonstrate that these melts have conductivities of hundreds of siemens per metre, much higher than CO_2-free hydrated melts or hydrated minerals. Our modelling, despite unavoidable simplifications, incorporates geochemical and petrological constraints and indicates that mantle with small fractions of CO_2-rich melts at 50–150 km reproduces both the electrical properties and the depth of the LAB quite well, whereas CO_2-free systems yield too poor or no agreement with geophysical observations. The presence of CO_2-rich incipient melts in the asthenosphere has important implications for radiogenic heat production because such melts are enriched in heat-producing elements such as potassium, uranium and thorium[30]. Moreover, the involvement of CO_2-rich incipient melts is also recognized in petrological processes occurring in the continental and in the cratonic LAB[30]. The association between the asthenosphere and incipient melting that we suggest here can therefore be extended to geodynamic settings other than the oceanic domains. However, it remains to be defined how the mechanical strength of the asthenosphere can be affected by small amounts of CO_2- and H_2O-rich melts and how this can be connected to tectonic plate motion.

METHODS SUMMARY

Starting materials were mixtures of basaltic glass and hydrated calcium and magnesium carbonates, such that H_2O and CO_2 were introduced into our sample with a constant molar CO_2/H_2O ratio of 2 (Extended Table 1a). Two dry carbonate melts were also investigated. The extent of dehydration and decarbonation of our samples during high-temperature, high-pressure experiments was shown to be small (compare Extended Table 1a with Extended Table 1b). The effect of H_2O was distinguished from that of CO_2 by using published experimental conductivity measurements on hydrated basalts[7] and the empirical equation (1). This empirical simplification is in line with the effect of H_2O on the conductivity of carbonated melts that we determined here (compare circle data points with square data points in Extended Data Fig. 3a): it is moderate and almost similar to that recently determined for basalts[7]. For simulations in Figs 2 and 3, equation (1) was used within a range of H_2O and CO_2 contents that required no extrapolation from our experiments.

Figures 2 and 3 were constructed on the assumption that incipient melts are well connected[20,24]. Reference 24 is the only work specifically tackling the connectivity of incipient carbonated melts in olivine, and it shows interconnection at the small melt fractions we are considering. Reference 20 confirmed the good connection for melt contents of ~1%. Mixtures of melt tubes and melt films have been considered

in our plots; the difference between these geometries implies conductivities differing by about 0.2 log units.

The isotherms in Fig. 4 were obtained using a model of sudden half-space cooling with similar parameters to those in ref. 8. The use of more complicated models, such as the plate model (see ref. 17, where, unlike here, the thickness of the lithosphere is decided a priori), yields similar isotherms[8] for the moderate depths we discuss here.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 October 2013; accepted 6 March 2014.

1. Höink, T., Jellinek, A. M. & Lenardic, A. Viscous coupling at the lithosphere-asthenosphere boundary. *Geochem. Geophys. Geosyst.* **12,** Q0AK02 (2011).
2. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144,** 93–108 (1996).
3. Karato, S. On the origin of the asthenosphere. *Earth Planet. Sci. Lett.* **321–322,** 95–103 (2012).
4. Evans, R. L. *et al.* Geophysical evidence from the MELT area for compositional controls on oceanic plates. *Nature* **437,** 249–252 (2005).
5. Naif, S., Key, K., Constable, S. & Evans, R. L. Melt-rich channel observed at the lithosphere-asthenosphere boundary. *Nature* **495,** 356–359 (2013).
6. Caricchi, L., Gaillard, F., Mecklenburgh, J. & Le Trong, E. Experimental determination of electrical conductivity during deformation of melt-bearing olivine aggregates: Implications for electrical anisotropy in the oceanic low velocity zone. *Earth Planet. Sci. Lett.* **302,** 81–94 (2011).
7. Ni, H., Keppler, H. & Behrens, H. Electrical conductivity of hydrous basaltic melts: implications for partial melting in the upper mantle. *Contrib. Mineral. Petrol.* **162,** 637–650 (2011).
8. Schmerr, N. The Gutenberg discontinuity: melt at the lithosphere-asthenosphere boundary. *Science* **335,** 1480–1483 (2012).
9. Hirschmann, M., Tenner, T., Aubaud, C. & Withers, A. C. Dehydration melting of nominally anhydrous mantle: the primacy of partitioning. *Phys. Earth Planet. Inter.* **176,** 54–68 (2009).
10. Green, D. H., Hibberson, W. O., Kovács, I. & Rosenthal, A. Water and its influence on the lithosphere-asthenosphere boundary. *Nature* **467,** 448–451 (2010).
11. Hier-Majumder, S. & Courtier, A. Seismic signature of small melt fraction atop the transition zone. *Earth Planet. Sci. Lett.* **308,** 334–342 (2011).
12. Faul, U. H. Melt retention and segregation beneath mid-ocean ridges. *Nature* **410,** 920–923 (2001).
13. Presnall, D. C. & Gudfinnsson, G. H. Carbonate-rich melts in the oceanic low-velocity zone and deep mantle. *Spec. Pap. Geol. Soc. Am.* **388,** 207–216 (2005).
14. Wallace, M. E. & Green, D. H. An experimental determination of primary carbonatite magma composition. *Nature* **335,** 343–346 (1988).
15. Green, D. H. & Liebermann, R. C. Phase-equilibra and elastic properties of a pyrolite model for oceanic upper mantle. *Tectonophysics* **32,** 61–92 (1976).
16. Fischer, K. M., Ford, H. A., Abt, D. L. & Rychert, C. A. The lithosphere-asthenosphere boundary. *Annu. Rev. Earth Planet. Sci.* **38,** 551–575 (2010).
17. Hirschmann, M. M. Partial melt in the oceanic low velocity zone. *Phys. Earth Planet. Inter.* **179,** 60–71 (2010).
18. Utada, H. & Baba, K. Estimating the electrical conductivity of the melt phase of a partially molten asthenosphere from seafloor magnetotelluric sounding data. *Phys. Earth Planet. Inter.* **227,** 41–47 (2013).
19. Gaillard, F., Malki, M., Iacono-Marziano, G., Pichavant, M. & Scaillet, B. Carbonatite melts and electrical conductivity in the asthenosphere. *Science* **322,** 1363–1365 (2008).
20. Yoshino, T., Laumonier, M., McIsaac, E. & Katsura, T. Electrical conductivity of basaltic and carbonatite melt-bearing peridotites at high pressures: implications for melt distribution and melt fraction in the upper mantle. *Earth Planet. Sci. Lett.* **295,** 593–602 (2010).
21. Hirano, N. *et al.* Volcanism in response to plate flexure. *Science* **313,** 1426–1428 (2006).
22. Dasgupta, R. *et al.* Carbon-dioxide-rich silicate melt in the Earth's upper mantle. *Nature* **493,** 211–215 (2013).
23. Stagno, V., Ojwang, D. O., McCammon, C. A. & Frost, D. J. The oxidation state of the mantle and the extraction of carbon from Earth's interior. *Nature* **493,** 84–88 (2013).
24. Minarik, W. G. & Watson, E. B. Interconnectivity of carbonate melt at low melt fraction. *Earth Planet. Sci. Lett.* **133,** 423–437 (1995).
25. Jones, A. G., Fullea, J., Evans, R. L. & Muller, M. R. Water in cratonic lithosphere: calibrating laboratory determined models of electrical conductivity of mantle minerals using geophysical and petrological observations. *Geochem. Geophys. Geosyst.* **13,** Q06010 (2012).
26. Cartigny, P., Pineau, F., Aubaud, C. & Javoy, M. Towards a consistent mantle carbon flux estimate: insights from volatile systematics (H_2O/Ce, δD, CO_2/Nb) in the North Atlantic mantle (14° N and 34° N). *Earth Planet. Sci. Lett.* **265,** 672–685 (2008).
27. Marty, B. The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth Planet. Sci. Lett.* **313–314,** 56–66 (2012).
28. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298,** 1–13 (2010).
29. Lizarralde, D., Chave, A., Hirth, G. & Schultz, A. Northeastern Pacific mantle conductivity profile from long-period magnetotelluric sounding using Hawaii-to-California submarine cable data. *J. Geophys. Res.* **100** (B9), 17837–17854 (1995).
30. O'Reilly, S. Y. & Griffin, W. L. The continental lithosphere-asthenosphere boundary: can we sample it? *Lithos* **120,** 1–13 (2010).

Acknowledgements This work, part of the ElectroLith project, benefited from funding by the European Research Council (ERC project #279790) and the French agency for research (ANR project #2010 BLAN62101). S.H.-M. acknowledges support from the US NSF grant EAR1215800 and a grant from the University of Orleans. We thank David H. Green for comments.

Author Contributions F.G. led the project and wrote the first draft. All authors contributed equally to the writing of subsequent drafts. D.S. and F.G. developed the experimental set-up, and D.S. performed the conductivity measurements. S.H.-M. contributed to the discussion and provided editorial assistance with the manuscript. D.S. and L.H. produced Fig. 1, E.G. and L.H. produced Fig. 2, D.S. produced Fig. 3, and L.H. and M.M. produced Fig. 4.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.G. (fabrice.gaillard@cnrs-orleans.fr).

METHODS

Starting materials. Electrical measurements were performed on six mixtures: two dry carbonated melts (CO_2 = 44–48 wt%), one hydrous carbonated melt (CO_2 = 25.9 wt%; H_2O = 10.2 wt%) and three hydrous carbonated basalts (CO_2 = 10.39–23.32 wt%; H_2O = 4.43–9.22 wt%). Starting materials used to obtain these mixtures were natural dolomite ($MgCa(CO_3)_2$), a natural basalt (popping rock[31]), salt (NaCl), sodium carbonate (Na_2CO_3) and brucite ($Mg(OH)_2$). See Extended Data Table 1a.

Experiments. All experiments were performed at 3 GPa in piston–cylinder apparatus (using half-inch graphite–Pyrex–talc assemblages), which were connected to a 1260 Solartron Impedance/Gain-Phase Analyzer for electrical conductivity measurements. The temperature was measured with a B-type thermocouple localized on top of the sample (Extended Data Fig. 1a). Oxygen fugacity (f_{O_2}) was not controlled during the measurements, but the presence of graphite (furnace) and molten carbonates (sample) should imply an oxygen fugacity close to FMQ-2 (ref. 23).

We developed a new protocol specifically adapted for electrical conductivity measurements on highly conductive molten materials (Extended Data Fig. 1). The new design employs a pseudo-four-wire configuration, which removes the electrical contribution of the electrical cell itself (Extended Data Fig. 2a). Such a configuration, previously used at 1 atm (refs 19, 32), is necessary for our measurements at pressure.

Cold pressed pellets (5-mm outer diameter) were cored to insert an inner Pt electrode (1 mm). A Pt foil surrounding the sample was used as outer electrode. An alumina jacket isolated the entire electrical cell from the graphite furnace. The sample impedance was measured between the two electrodes arranged in a coaxial geometry[33,34]. The inner electrode was connected to the impedance spectrometer via the two wires of the thermocouple[34]. The outer electrode was connected to a nickel cylinder (located 5 mm above the sample) that was mounted in series with two additional wires (B-type thermocouples) (Extended Data Fig. 1a).

Impedance spectra, conductivity calculations and uncertainties. The impedance spectra were collected during heating and cooling cycles (Extended Data Table 2a) at different temperature plateaux in the frequency range 1 Hz–1 MHz. Unlike the spectra collected at low temperatures (that is, solid samples showing impedance arcs), high-temperature spectra (that is, molten samples) had vertical lines (Extended Data Fig. 2b). These spectra correspond to inductance-dominated signals, and the intercept of each spectrum with the *x* axis yielded the resistance of the sample.

Reproducibility of electrical measurements was validated by performing the measurements during heating and cooling cycles. For the experiments on hydrated samples (HC and HCB-9, -7, -4), a step of about 10 min was performed at 700 °C (before brucite dehydration), and the temperature was then rapidly raised (<10 s) to 1,300–1,410 °C (that is, the temperature of the molten state), which limited sample dehydration.

Data reductions and uncertainties. The electrical conductivities of the samples were calculated from the measured resistances using the following relationship[33,34]:

$$\sigma = \frac{\ln(r_{out}/r_{in})}{2\pi h R} \qquad (2)$$

with σ being the electrical conductivity in S m^{-1}; r_{out}, r_{in} and h respectively being the outer radius, the inner radius and the height of the samples in metres; and R being the resistance of the sample in Ω (Extended Data Figs 1b and 2b).

Uncertainties in σ were calculated considering geometrical factors of the samples (Extended Data Fig. 1b) and propagated errors of each measured resistance. The uncertainties in σ are 7% on average for all measurements and reach a maximum of 16% in HCB-4.

Sample characterization. Scanning electron microscope (SEM) imaging and electron microprobe analyses (EMPAs) were systematically performed after each experiment. Determination of r_{out}, r_{in} and h by SEM imaging showed an average decrease of 20% compared with the initial geometry, most probably due to porosity loss during melting (Extended Data Fig. 1b). No melt leak was observed, and the entire sample remained sandwiched between the MgO plugs and the electrodes.

EMPAs were conducted at 15 keV and 10 nA, with 10-s counting on peak elements. The beam size (100 μm × 100 μm) was adapted to obtain average chemical compositions, smoothing the heterogeneities due to quench crystallizations. Compositions before and after experiments indicate no contamination by the MgO surrounding the sample and no considerable volatile loss from the sample (Extended Data Table 1b). Carbon dioxide contents were determined using the by-difference method[22] and indicate negligible decarbonation.

A Flash 2000 elemental analyser (Thermo Scientific) was used to measure the H_2O contents of samples before and after experiments. Samples were heated to >1,500 °C and the released H_2O was reduced into elemental H, which was detected by a highly sensitive thermal conductivity detector. This gave the H2O content with a precision of ±0.5 wt%. We observed negligible dehydration during conductivity measurements.

Conductivity results. Extended Data Fig. 3a shows the good reproducibility of the electrical measurements during heating and cooling cycles. The conductivity–temperature relationship for each sample was fitted using an Arrhenius law

$$\sigma = \sigma_0 e^{-E_a/RT}$$

with σ_0, E_a, R and T as define in equation (1).

Calculated pre-exponential factors and activation energies are presented in Extended Data Table 2b.

Increasing the CO_2 concentration in the melt drastically increases its conductivity. Furthermore, we observed that CO_2 tends to decrease both activation energies and pre-exponential factors.

Conductivity modelling. The semi-empirical law that we have developed can be considered the sum of two conductive processes operating in carbonated basalts: conduction in the hydrous silicate melts by interstitial sodium mobility[7,33,34] and conduction in CO_2-rich melts caused by the motion of all species in ionic liquids[19] (equation (1)).

The pre-exponential factor, σ_0, and the activation energy, E_a, for both H_2O and CO_2 are related by a compensation law[35,36] (Extended Data Fig. 3b):

$$\ln(\sigma_0^{\text{volatile}}) = dE_A^{\text{volatile}} + e \qquad (3)$$

The decrease in activation energy as a function of volatile content is exponential:

$$E_A^{\text{volatile}} = a\exp\left(-bC^{\text{volatile}}\right) + c \qquad (4)$$

C^{volatile} is the CO_2 or H_2O content in wt%.

Thus, equations (1), (3) and (4) directly relate melt conductivities to melt H_2O and CO_2 contents. We have determined the Arrhenius parameters for the melt composition as a function of H_2O content using data in ref. 7 in the temperature range 1,200–1,500 °C. Parameters *a*, *b*, *c*, *d* and *e* for H_2O were obtained by fitting these data with equations (3) and (4) (Extended Data Table 3). The best parameters for CO_2 were obtained by minimizing the differences between our measured conductivities and the CO_2-free values returned by equation (1) (Extended Data Table 3). In doing so, we assumed that the effect of H_2O on electrical conductivity is similar and modest in both silicate and carbonate melts, which is indeed what our measurements show (compare samples C and HC in Extended Data Fig. 3a). Our model reproduces the experimental measurements of σ (ours and that of ref. 7) with an average error of 5% (maximum 10%).

Figures 2 and 3. The bulk rock is considered a peridotite containing a fraction of interconnected melt, where volatiles partition between the solid and the melt phase. The bulk H_2O content is related to H_2O in melt and in peridotite by

$$C_{H_2O}^{\text{melt}} = \frac{C_{H_2O}^{\text{bulk}}}{X_m^{\text{melt}} + \left(1 - X_m^{\text{melt}}\right) D_{H_2O}^{\text{perid/melt}}}$$

$$C_{H_2O}^{\text{perid}} = \frac{C_{H_2O}^{\text{bulk}} D_{H_2O}^{\text{perid/melt}}}{X_m^{\text{melt}} + \left(1 - X_m^{\text{melt}}\right) D_{H_2O}^{\text{perid/melt}}}$$

where X_m^{melt} is the mass fraction of melt and $D_{H_2O}^{\text{perid/melt}}$ is the partition coefficient of H_2O between peridotite and melt (0.007, the average partition coefficient over the pressure range 1.5–4 GPa; refs 9, 37).

The concentration of H_2O in olivine is

$$C_{H_2O}^{\text{ol}} = C_{H_2O}^{\text{perid}} \frac{D_{H_2O}^{\text{ol/melt}}}{D_{H_2O}^{\text{perid/melt}}}$$

which differs from that of peridotite because $D_{H_2O}^{\text{ol/melt}}$ is about 0.002 (ref. 9). (See also fig. 6 of ref. 9 for the modal proportion of mineral phases in peridotite $Ol_{58}Opx_{28}Cpx_{12}Spn_2$.)

Pargasite amphibole was considered to affect the distribution of H_2O for bulk H_2O content exceeding 200 p.p.m. (enriched mantle in Figs 2c and 3 and Extended Data Fig. 4). We computed that the H_2O content exceeding 200 p.p.m. in bulk goes in pargasite and computed the partitioning of H_2O among NAMs following ref. 9. If temperatures exceed ~1,070 °C (Extended Data Fig. 5), then pargasite dehydration melting occurs and H_2O partitions between melt and the solids as described above.

CO_2 distributes exclusively in the liquid phase[23,38]; that is, $D_{CO_2}^{\text{perid/melt}} = 0$. Therefore

$$C_{CO_2}^{\text{melt}} = \frac{C_{CO_2}^{\text{bulk}}}{X_m^{\text{melt}}}$$

$$C_{CO_2}^{\text{perid}} = 0$$

At small melt fractions, this can lead to CO_2 concentrations higher than that of carbonate (45 wt%) or, in other words, to CO_2 saturation. Calculations performed at saturated concentrations are mentioned in Fig. 2b, c.

Conversion from mass to volume fraction of melt is done considering volume properties of silicate melts[39] and carbonate melts[40,41]:

$$X_{v}^{melt}=\left(1+\left(\frac{1}{X_{m}^{melt}}-1\right)\frac{d_{melt}}{d_{perid}}\right)^{-1}$$

Here d_{perid} is the density of peridotite, that is, 3.3, and d_{melt} is the density of melt. Because there are no data for the density of hydrous carbonated basalts, we estimate the density of these melts using a simple mixing law

$$d_{melt}=\frac{C_{H_2O}^{melt}}{100}d_{H_2O}+\frac{C_{carb}^{melt}}{100}d_{carb}+\left(1-\frac{C_{H_2O}^{melt}+C_{carb}^{melt}}{100}\right)d_{basalt}$$

where concentrations are expressed as wt% and $d_{H_2O}=1.4$ (considering 12 cm^3 mol^{-1} for H_2O partial volume in melts). The mass per cent of carbonate (M-CO_3, where M denotes cations such as Ca, Mg and others) in melt, C_{carb}^{melt}, is approximated as $2C_{CO_2}^{melt}$. Carbonate density, d_{carb}, was set to 2.4 according to the density of molten carbonates at 1 bar (ref. 40). The use of the CO_2 partial molar volume calculated in ref. 41 yields similar density results. The density of basalt, d_{basalt}, was taken to be 2.8 (ref. 39).

The conductivity of the melt, $\sigma_{melt}(C_{H_2O}^{melt},C_{CO_2}^{melt},T)$, was calculated using equation (1). The conductivity of the peridotite was assumed to be controlled by that of hydrous olivine[25].

The bulk conductivity was calculated using the mean of tube[42–44]

$$\sigma_{tube}=\frac{1}{3}X_{v}^{melt}\sigma_{melt}+\left(1-X_{v}^{melt}\right)\sigma_{ol}$$

and film[44–47]

$$\begin{aligned}\sigma_{layer}=&\left\{\left[\left(1-X_{v}^{melt}\right)^{2/3}-1\right]\sigma_{melt}-(1-X_{v}^{melt})^{1/3}\sigma_{ol}\right\}\\&\times\left\{\left[1-X_{v}^{melt}-\left(1-X_{v}^{melt}\right)^{2/3}\right]\sigma_{melt}\right.\\&\left.+\left[\left(1-X_{v}^{melt}\right)^{2/3}-\left(1-X_{v}^{melt}\right)-1\right]\sigma_{ol}\right\}^{-1}\sigma_{melt}\end{aligned}$$

Remarkably, for the fraction of liquid considered in this study, the difference between the tube and layer conductivities is mostly lower than 0.2 log units. We therefore averaged the two expressions:

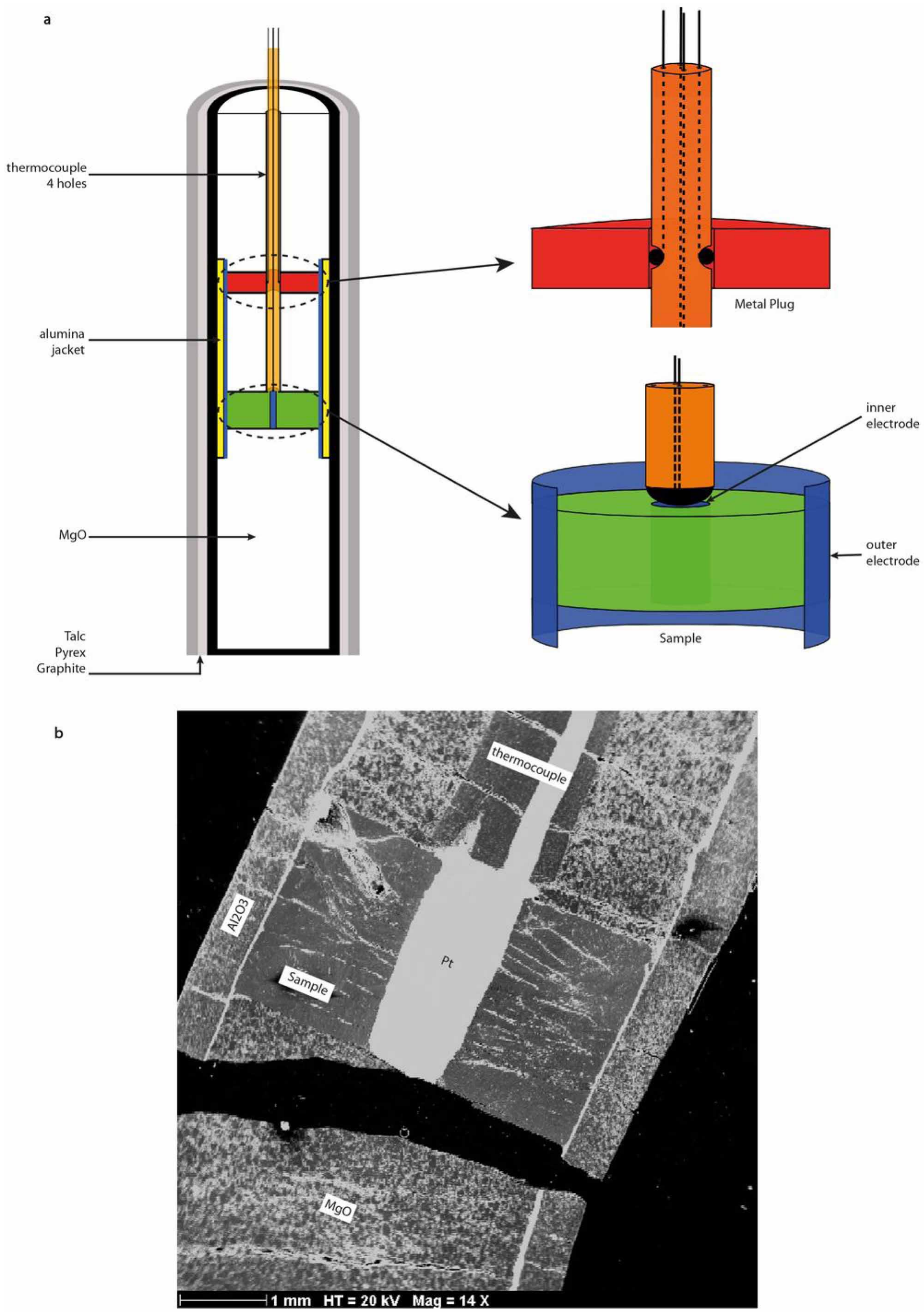
$$\sigma_{bulk}\left(X_{v}^{melt},C_{H_2O}^{bulk},C_{CO_2}^{bulk},T\right)=\frac{\sigma_{tube}+\sigma_{layer}}{2}\qquad(5)$$

This equation returns values almost the same as those reported in refs 20, 46. The bulk conductivity of partially molten peridotite reported in Figs 2 and 3, as well as in the text, was calculated using equation (5).

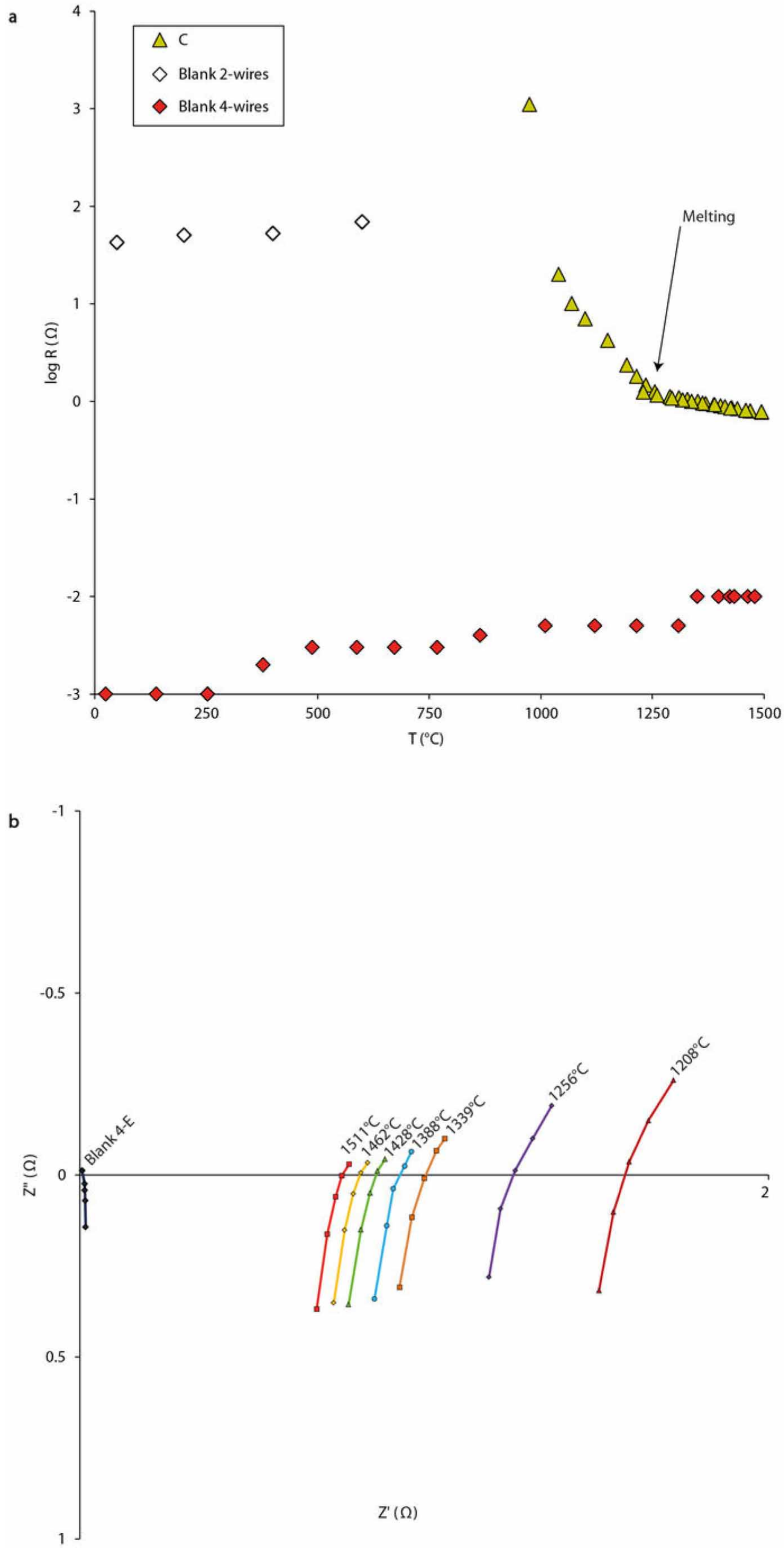
The melt fraction in Fig. 3 was approximated as[48] $X_{m}^{melt}=2.5CO_2^{bulk}+6H_2O^{bulk}$ (Extended Data Fig. 6).

Buoyant basalts versus incipient melts. An impermeable layer has been suggested to prevent the melt prevailing in the LAB from rising to the surface[49]. The rate of melt ascent due to buoyancy is otherwise expected to be of the order of several cm yr^{-1} (refs 11, 12) if melt content is 3–5 vol%. Our model of incipient melting implies an impermeable boundary that is caused by phase relationships[14], that is, a thermodynamic boundary through which melt cannot rise. We furthermore emphasize the limited melt mobility[50] at the small melt fraction of interest because, in particular, surface tensions would unavoidably tend to retain the buoyant melt[51]. To conclude, if basalts, being anyway not thermodynamically stable in the asthenosphere, tend to migrate out of the asthenosphere[52], small melt fractions may in contrast be mechanically stable in the LAB.

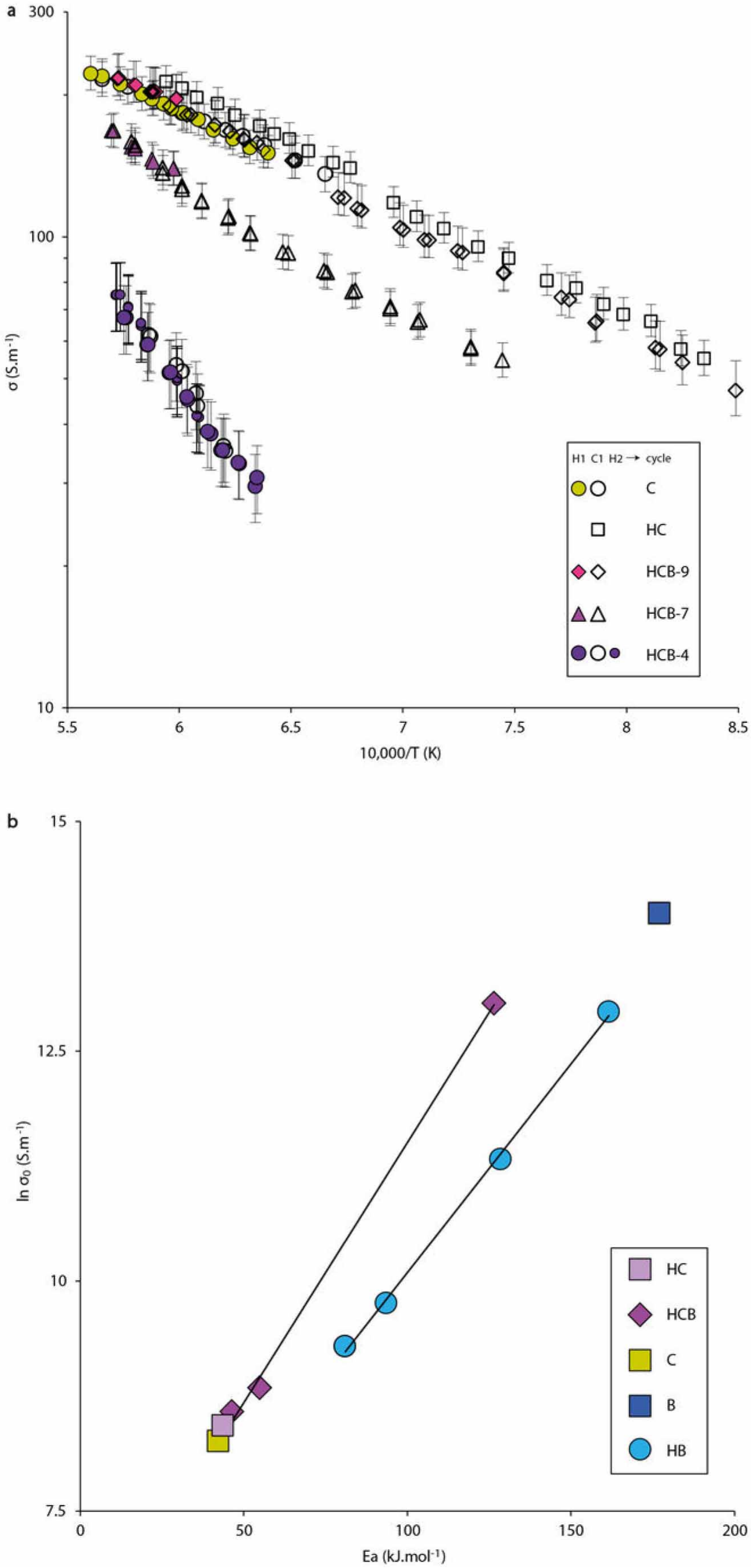
31. Javoy, M. & Pineau, F. The volatiles record of a "popping" rock from the Mid-Atlantic Ridge at 14° N: chemical and isotopic composition of gas trapped in the vesicles. *Earth Planet. Sci. Lett.* **107,** 598–611 (1991).
32. Pommier, A., Gaillard, F., Malki, M. & Pichavant, M. Methodological re-evaluation of the electrical conductivity of silicate melts. *Am. Mineral.* **95,** 284–291 (2010).
33. Gaillard, F. Laboratory measurements of electrical conductivity of hydrous and dry silicic melts under pressure. *Earth Planet. Sci. Lett.* **218,** 215–228 (2004).
34. Hashim, L. *et al.* Experimental assessment of the relationships between electrical resistivity, crustal melting and strain localization beneath the Himalayan-Tibetan belt. *Earth Planet. Sci. Lett.* **373,** 20–30 (2013).
35. Pommier, A., Gaillard, F., Pichavant, M. & Scaillet, B. Laboratory measurements of electrical conductivities of hydrous and dry Mount Vesuvius melts under pressure. *J. Geophys. Res. Solid Earth* **113,** B05205 (2008).
36. Tyburczy, J. & Waff, H. S. Electrical conductivity of molten basalt and andesite to 25 kilobars pressure: geophysical significance and implications for the charge transport and melt structure. *J. Geophys. Res.* **88,** 2413–2430 (1983).
37. Katz, R. F., Spiegelman, M. & Langmuir, C. H. A new parameterization of hydrous mantle melting. *Geochem. Geophys. Geosyst.* **4,** 1073 (2003).
38. Keppler, H., Wiedenbeck, M. & Shcheka, S. S. Carbon solubility in olivine and the mode of carbon storage in the Earth's mantle. *Nature* **424,** 414–416 (2003).
39. Lange, R. A. & Carmichael, I. S. E. Thermodynamic properties of silicate liquids with emphasis on density thermal expansion and compressibility. *Rev. Mineral.* **24,** 25–64 (1990).
40. Liu, Q. & Lange, R. A. New density measurements on carbonate liquids and the partial molar volume of the $CaCO_3$ component. *Contrib. Mineral. Petrol.* **146,** 370–381 (2003).
41. Guillot, B. & Sator, N. Carbon dioxide in silicate melts: a molecular dynamics simulation study. *Geochim. Cosmochim. Acta* **75,** 1829–1857 (2011).
42. ten Grotenhuis, S. M., Drury, M. R., Spiers, C. J. & Peach, C. J. Melt distribution in olivine rocks based on electrical conductivity measurements. *J. Geophys. Res.* **110,** B12201 (2005).
43. Hammouda, T. & Laporte, D. Ultrafast mantle impregnation by carbonatite melts. *Geology* **28,** 283–285 (2000).
44. Glover, P. W. J., Hole, M. J. & Pous, J. A modified Archie's law for two-conducting phases. *Earth Planet. Sci. Lett.* **180,** 369–383 (2000).
45. Partzsch, G. M., Schilling, F. R. & Arndt, J. The influence of partial melting on the electrical behaviour of crustal rocks: laboratory examinations, model calculations and geological interpretations. *Tectonophysics* **317,** 189–203 (2000).
46. Yoshino, T., McIsaac, E., Laumonier, M. & Katsura, T. Electrical conductivity of partial molten carbonate peridotite. *Phys. Earth Planet. Inter.* **194–195,** 1–9 (2012).
47. Garapić, G., Faul, U. H. & Brisson, E. High-resolution imaging of the melt distribution in partially molten upper mantle rocks: evidence for wetted two-grain boundaries. *Geochem. Geophys. Geosyst.* **14,** 556–566 (2013).
48. Green, D. H. & Falloon, T. J. Primary magmas at mid-ocean ridges, "hotspots," and other intraplate settings: constraints on mantle potential temperature. *Spec. Pap. Geol. Soc. Am.* **388,** 217–247 (2005).
49. Katz, R. F. & Weatherley, S. M. Consequences of mantle heterogeneity for melt extraction at mid-ocean ridges. *Earth Planet. Sci. Lett.* **335-336,** 226–237 (2012).
50. Hier-Majumder, S., Ricard, Y. & Bercovici, D. Role of grain boundaries in magma migration and storage. *Earth Planet. Sci. Lett.* **248,** 735–749 (2006).
51. Takei, Y. & Holtzman, B. K. Viscous constitutive relations of solid-liquid composites in terms of grain boundary contiguity: 1. Grain boundary diffusion control model. *J. Geophys. Res.* **114,** B06205 (2009).
52. Sakamaki, T. *et al.* Ponded melt at the boundary between the lithosphere and asthenosphere. *Nat. Geosci.* **6,** 1041–1044 (2013).
53. Hirschmann, M. M. Mantle solidus: experimental constraints and the effects of peridotite composition. *Geochem. Geophys. Geosyst.* **1,** 1042 (2000).
54. Dasgupta, R. Ingassing, storage, and outgassing of terrestrial carbon through geologic time. *Rev. Mineral. Geochem.* **75,** 183–229 (2013).
55. Iacono-Marziano, G., Morizet, Y., Le Trong, E. & Gaillard, F. New experimental data and semi-empirical parameterization of H_2O-CO_2 solubility in mafic melts. *Geochim. Cosmochim. Acta* **97,** 1–23 (2012).



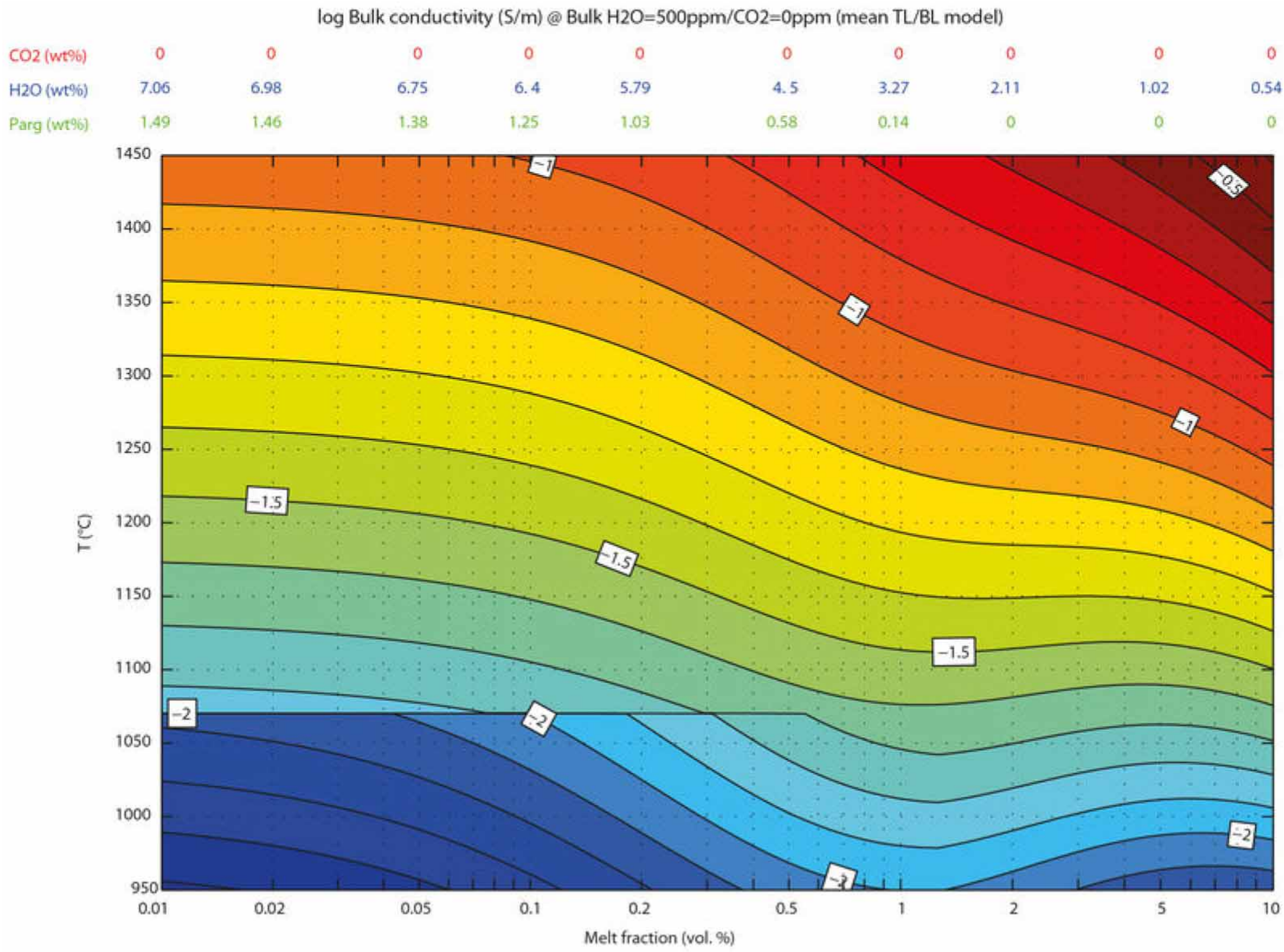
Extended Data Figure 1 | Set-up of electrical conductivity measurement using four wires. **a**, Modified piston–cylinder assembly for electrical conductivity measurements using a four-wire configuration. The cored sample (in green) contains in its centre an inner electrode in platinum (in blue). A platinum foil (in blue) surrounds the sample, which extends upwards and downwards from the sample and corresponds to the outer electrode. The sample is sandwiched by machined MgO ceramics (in white). The electrode-sample assemblage is isolated from the graphite furnace by an Al_2O_3 jacket (in yellow). The four-electrode wires are emplaced using a four-hole Al_2O_3 tube (in orange). Two of these wires, that is, the thermocouple, are in contact with the inner electrode, whereas the outer electrode is in contact with two other wires by means of a top Ni plug (in red). **b**, SEM image of the assemblage of sample C after experiments (up to 1,463 °C and 3 GPa). We observed an average decrease of 20% compared to the initial cell geometry (corresponding to the porosity loss during melting). Cell geometry parameters (h, r_{in} and r_{out} in equation (2)) are determined from SEM images for each sample.



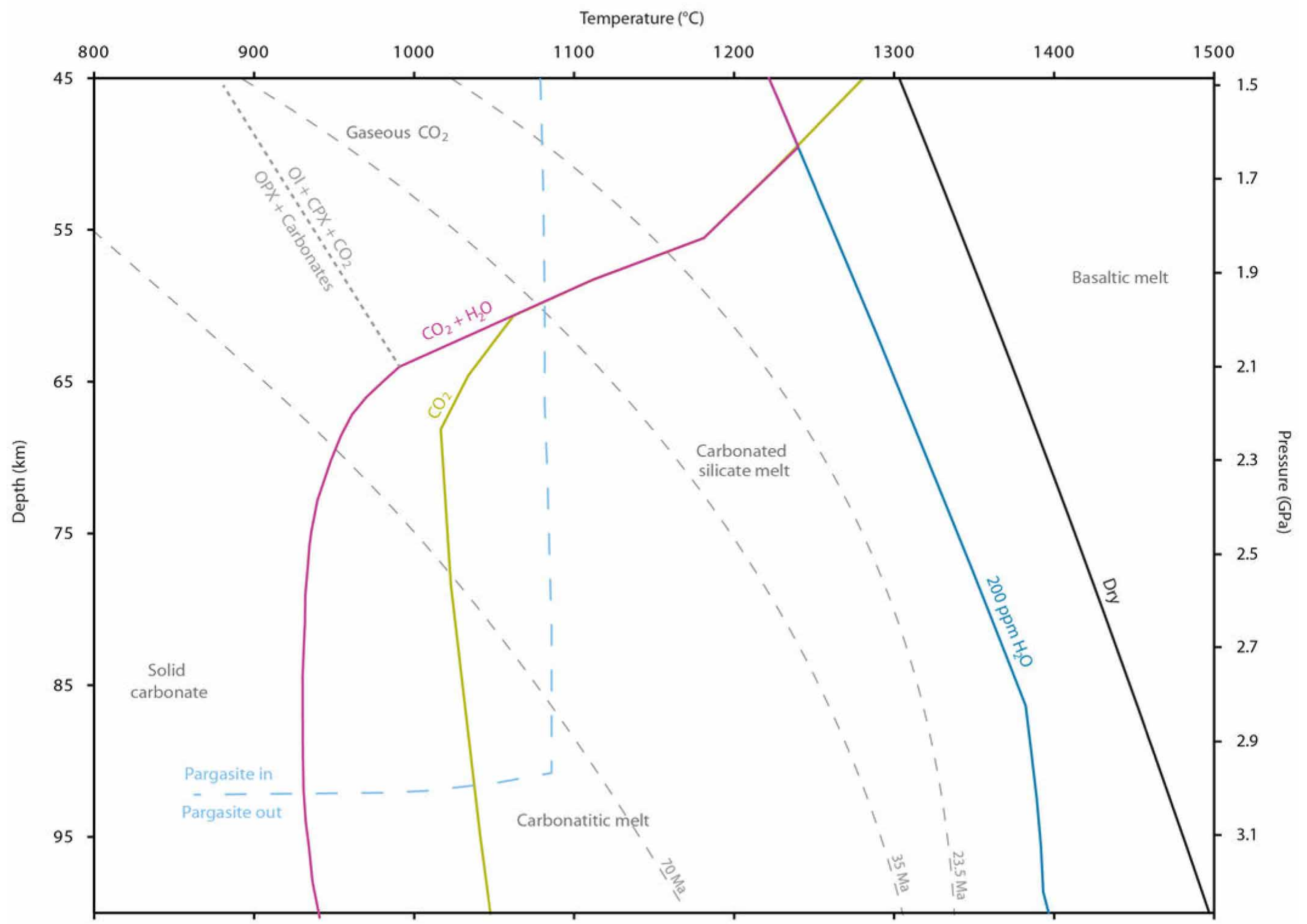
Extended Data Figure 2 | Measured resistance of molten carbonate versus nickel. **a**, The electrical cell resistance versus temperature. We show the resistance of a sample made of nickel measured using either a two-wire set-up (empty diamond) or a four-wire set-up (red diamond). There are several orders of magnitude of difference between the two measurements, showing that the two-wire setup is not suitable at all for conductive materials. We also show the resistance of carbonate in a four-wire set-up (sample C, molten at $T > 1{,}230\,^{\circ}\mathrm{C}$; green triangle). **b**, Impedance spectra obtained on molten carbonate (sample C) at 3 GPa as a function of temperature. Impedance spectra show vertical lines, indicating an inductance-dominated signal for all temperatures. The resistance is taken from the intercept with the horizontal axis. Data are obtained at frequencies ranging from 19,905 to 315,479 Hz. The black line represents an impedance spectrum of a nickel sample (blank) obtained with a four-wire configuration at $1{,}464\,^{\circ}\mathrm{C}$.



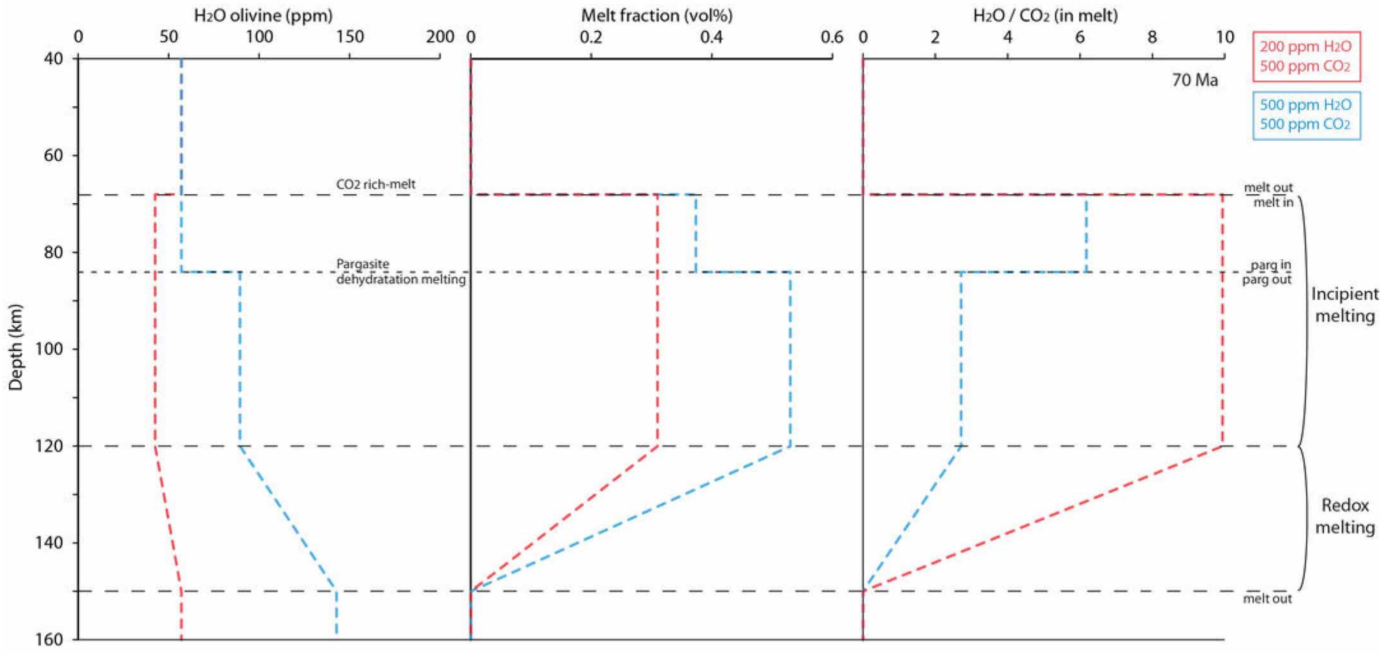
Extended Data Figure 3 | Electrical conductivity measurements. **a**, Electrical conductivity versus reciprocal temperature measured on carbonated melts and hydrous carbonated basalts. Samples: a carbonated melt (C), a hydrous carbonated melt (HC), and three hydrous carbonated basalts with H_2O contents ranging from 4.43 to 9.22 wt% (HCB-9, HCB-7 and HCB-4) and CO_2 contents ranging from 10.39 to 23.32 wt%. To complete Fig. 1, we distinguished heating–cooling temperature cycles and reported error bars. Large solid symbols, heating cycle (H1); open symbols, cooling cycle (C1); small solid symbols, second heating cycle (H2) (compare with Extended Data Table 2a). The error bars include uncertainties in the geometrical factors of the samples and in the measured resistance. **b**, Compensation plots showing the correlation between activation energy, E_a, and pre-exponential terms, $\ln(\sigma_0)$. Hydrous basalts (HB) are from the experimental data set of ref. 7 between 1,200 and 1,500 °C, and the data point for the dry basalt (B) is from ref. 32. The dry carbonated melt (C), the hydrous carbonated melts (HC) and the hydrous carbonated basalts (HCB) are from this study (see Extended Data Table 2b for the Arrhenius parameters).



Extended Data Figure 4 | The incipient melt effect on the electrical conductivity of an H_2O-enriched, CO_2-free peridotite. This figure completes the scenarios illustrated in Fig. 2. The conductivity of partially molten peridotite, in which H_2O partitions between minerals and melt (Methods), is reported as a function of melt content and temperature for CO_2-free peridotite with 500 p.p.m. H_2O (log values; conductivity increases from cold to warm colours). The discontinuity at $T = 1{,}070$ °C is due to pargasite amphibole breakdown (Extended Data Fig. 5) that redistribute H_2O between NAMs and the melt as explained in the Methods. Melt H_2O contents (blue if pargasite out, green if pargasite in) are tabulated above the panel.



Extended Data Figure 5 | Melting curves for different bulk peridotitic systems as functions of temperature and depth. The solidus of dry peridotite (black curve) is calculated from ref. 53. The dehydration solidus of nominally anhydrous peridotite at 200 p.p.m. H_2O (blue curve) is modelled from ref. 9. The dehydration solidus of pargasite lherzolite is based on ref. 10. The nominally anhydrous carbonated, fertile peridotite solidus is based on ref. 54 and references therein (green curve). The H_2O-undersaturated carbonated, fertile peridotite curve (purple curve) corresponds to the solidus of a pyrolite with 0.5–2.5 wt% CO_2 and 0.3 wt% H_2O (ref. 14). For pressures $\leq$1.7 GPa, carbonated melts are unstable and gaseous CO_2 prevails. We connected the melting curve of CO_2-bearing peridotite to that of the dry peridotite at low pressures, which slightly differs from previously published phase diagrams. We considered that, for $P \leq 1.7$ GPa, gaseous CO_2 must have a negligible influence on the peridotite solidus due to the small solubility of CO_2 in basaltic melts[55]. Similarly, at low pressures, the H_2O-undersaturated carbonated, fertile peridotite solidus was connected to the dehydration solidus of nominally anhydrous peridotite (considering peridotite with 200 p.p.m. H_2O), neglecting the presence of pargasite owing to the NAM's H_2O capacity storage.



Extended Data Figure 6 | Phase equilibria control on H_2O–CO_2 partitioning, ultimately resulting in a change in conductivity as shown in Fig. 3. We show changes in H_2O content in olivine (left), melt fraction (centre) and melt CO_2/H_2O (right) for the 70-Myr age used for calculation in Fig. 3. We use two illustrative compositions: bulk with 200 p.p.m. H_2O and 500 p.p.m. CO_2 and bulk with 500 p.p.m. H_2O and 500 p.p.m. CO_2.

Extended Data Table 1 | Chemical composition of samples before and after electrical conductivity measurements

a

	Carbonated Melt			Hydrous Carbonated Basalt			Starting Material	
	dry		hydrous					
Sample	C	C +	H C	H C B -9	H C B -7	H C B -4	Dolomite	Basalt
SiO_2	0.24	0.24	0.02	5.23	15.67	31.32	0.27	52.18
TiO_2	0.03	0.03	0.02	0.20	0.56	1.10	0.03	1.81
Al_2O_3	0.02	0.02	0.01	1.52	4.54	9.07	0.02	15.11
FeO	0.20	0.20	0.02	0.82	2.41	4.80	0.22	7.99
MgO	18.68	18.68	23.79	22.16	18.90	14.01	20.76	7.49
CaO	26.50	26.50	32.19	30.01	25.66	19.13	29.44	10.42
Na_2O	0.02	5.87	0.01	0.31	0.91	1.82	0.02	3.02
K_2O	0.03	0.03	0.01	0.07	0.21	0.41	0.03	0.67
MnO	0.03	0.03	0.01	0.03	0.06	0.11	0.03	0.17
P_2O_5	0.33	0.33	0.33	0.36	0.38	0.41	0.37	0.49
NaCl	10.00	-	7.50	6.75	5.25	3.00	-	-
Total	56.07	51.92	63.91	67.46	74.55	85.18	51.19	0.00
CO_2	43.93	48.08	25.91	23.32	18.15	10.39	48.81	0.05
H_2O	-	-	10.18	9.22	7.30	4.43	-	0.60

b

	Carbonated Melt			Hydrous Carbonated Basalt		
	dry		hydrous			
Sample	C	C +	H C	H C B -9	H C B -7	H C B -4
SiO_2	0.4	0.32	0.26	5.49	15.83	32.12
TiO_2	0.03	0.03	0.01	0.29	0.22	1.05
Al_2O_3	0.07	0.02	0.35	1.25	4.82	10.43
FeO	0.01	0.03	0.03	0.04	0.29	3.98
MgO	17.76	18.02	23.39	22.94	20.69	14.52
CaO	26.63	26.74	32.05	30.29	26.16	19.25
Na_2O	5.32	5.65	3.75	3.95	3.54	3.02
K_2O	0.05	0.03	0.01	0.1	0.05	0.48
MnO	0.04	0.03	0.05	0.08	0.03	0.1
P_2O_5	0.33	0.33	0.04	0.38	0.27	0.51
NaCl	6.12	-	4.51	4.02	3.11	1.92
Total	56.76	51.19	64.45	68.83	75.01	87.38
CO_2	-	-	10.11	8.09	6.89	3.46
H_2O	43.24	48.81	25.44	23.08	18.10	9.16

a, Bulk chemical composition of the starting materials (wt%). **b**, Analysed chemical compositions (wt%) of the bulk systems after each experimental runs. H_2O contents were analysed using an elemental analyser (Flash 2000). Carbon dioxide contents were obtained by difference on EMPA.

Extended Data Table 2 | Temperature range of electrical conductivity measurements and adjusted Arrhenius parameters

a

	C	C+	HC	HBC-9	HBC-7	HBC-4
Heating cycle	0-1511	0-1404	-	1397-1472	1383-1482	1304-1465
Cooling cycle	1511-1120	-	1410-650	1472-850	1482-1070	1465-1338
Heating cycle 2	-	-	-	-	-	1338-1470

b

	Ea ($J.mol^{-1}$)	error ($J.mol^{-1}$)	ln σ_0 ($S.m^{-1}$)	error ($S.m^{-1}$)
C	42150	700	8.26	0.05
C+	41950	1330	8.2	0.12
HC	43460	670	8.47	0.05
HBC-9	46360	2300	8.58	0.17
HBC-7	54900	3400	8.84	0.32
HBC-4	126500	2700	13.02	0.2

a, Heating and cooling cycles. Cycles of the different experimental runs with corresponding temperature ranges (°C). Runs HC and HCB-9, -7 and -4 were heated from room temperature to 700 °C. A step of 10 min was performed at 700 °C before a rapid increase to the measurement temperature. **b**, Arrhenius parameters and their errors determined for each melt studied.

Extended Data Table 3 | Parameters used for equations (3) and (4)

	H_2O	CO_2
a	88,774	789,166
b	0.3880	0.1808
c	73,029	32,820
d	4.54E-05	5.50E-05
e	5.5607	5.7956

LETTER

doi:10.1038/nature13265

Widespread decline of Congo rainforest greenness in the past decade

Liming Zhou[1], Yuhong Tian[2], Ranga B. Myneni[3], Philippe Ciais[4], Sassan Saatchi[5], Yi Y. Liu[6], Shilong Piao[7], Haishan Chen[8], Eric F. Vermote[9], Conghe Song[10,11] & Taehee Hwang[12]

Tropical forests are global epicentres of biodiversity and important modulators of climate change[1], and are mainly constrained by rainfall patterns[1–3]. The severe short-term droughts that occurred recently in Amazonia have drawn attention to the vulnerability of tropical forests to climatic disturbances[4–9]. The central African rainforests, the second-largest on Earth, have experienced a long-term drying trend[10,11] whose impacts on vegetation dynamics remain mostly unknown because *in situ* observations are very limited. The Congolese forest, with its drier conditions and higher percentage of semi-evergreen trees[12,13], may be more tolerant to short-term rainfall reduction than are wetter tropical forests[11], but for a long-term drought there may be critical thresholds of water availability below which higher-biomass, closed-canopy forests transition to more open, lower-biomass forests[1,2,14]. Here we present observational evidence for a widespread decline in forest greenness over the past decade based on analyses of satellite data (optical, thermal, microwave and gravity) from several independent sensors over the Congo basin. This decline in vegetation greenness, particularly in the northern Congolese forest, is generally consistent with decreases in rainfall, terrestrial water storage, water content in aboveground woody and leaf biomass, and the canopy backscatter anomaly caused by changes in structure and moisture in upper forest layers. It is also consistent with increases in photosynthetically active radiation and land surface temperature. These multiple lines of evidence indicate that this large-scale vegetation browning, or loss of photosynthetic capacity, may be partially attributable to the long-term drying trend. Our results suggest that a continued gradual decline of photosynthetic capacity and moisture content driven by the persistent drying trend could alter the composition and structure of the Congolese forest to favour the spread of drought-tolerant species[1,2,14].

The impact of changes in precipitation patterns, such as short-term and long-term droughts, on tropical rainforests is poorly understood and currently under debate[4–11]. Systematic monitoring of the forests is essential to understanding their response to climate change, and remote sensing remains the only viable way of synoptically and repeatedly monitoring vast remote regions such as the Congo basin[10,11]. This study uses Enhanced Vegetation Index (EVI)[15] data derived from a satellite-borne sensor, MODerate resolution Imaging Spectroradiometer (MODIS), for the period 2000–2012. EVI correlates well with leaf area index, canopy photosynthetic activity and primary productivity[16–18]. We focus our study on intact forested regions in the Congo basin (5° N–6° S, 14° E–31° E) during April–May–June, which represents the first of two rain and peak growing seasons and exhibits the highest percentage of forested area with high-quality EVI data (Extended Data Fig. 1). EVI lags the bimodal seasonal cycle of rainfall by about one month and has a smoother seasonal variation than rainfall, consistent with observed phenological (leaf area index) responses of tropical trees to increasing soil moisture[19].

We also use three gauge-measured and satellite-derived rainfall data sets[20–22] and other satellite products: terrestrial water storage (TWS)[23,24], aerosol optical thickness (AOT), cloud optical thickness (COT), photosynthetically active radiation (PAR) and land surface temperature (LST) as climate drivers; and vegetation optical depth (VOD)[25] and canopy backscatter anomaly (CBA)[11] (together with EVI) as vegetation variables (see Methods). VOD represents water content in aboveground woody and leaf biomass and is sensitive to long-term climate changes[25]. CBA reflects the changes in structure and moisture in upper forest layers and thus can help identify large-scale tree mortality[8,11]. TWS quantifies large-scale and low-frequency total ground, surface and vegetation water storage anomalies[23,24]. Unlike EVI, the microwave products CBA and VOD are least affected by atmospheric and weather conditions[11,25]. Most of the data are independent and thus allow a multi-factor analysis.

Although differing in data source, duration, spatial resolution and processing, the three rainfall data sets show strong and similar interannual variations during April–May–June over the study region, with the strongest negative anomalies falling in the last decade of the long-term 1950 to 2012 mean (Fig. 1a). The regional-mean rainfall declined significantly by -0.32 ± 0.10 mm per day per decade ($7.2 \pm 2.2\%$, $P = 0.002$) or by -0.56 mm per day (12.6%) between the last and first decades for the period 1985–2012. The drying trend (Fig. 1b and c) is widespread across the study region, with 25%–62% of forested area showing a significant negative trend ($P < 0.05$).

The spatial patterns of EVI trends are shown in Fig. 2, together with the corresponding trends in rainfall, TWS and CBA for the period 2000–2012. Because most of the satellite data sets are only 10–13 years long, linear regressions are used to quantify simply whether there is a trend within each data record; such a trend, however, cannot be extrapolated linearly over longer periods. Although the time series is short, EVI declined over 92% of the study area from 2000 to 2012 and in 97% of the area from 2003 to 2012, with 39% and 54% of the area showing a significant negative trend ($P < 0.1$), respectively, indicating that the EVI decrease became broader in space and stronger over time. The two rainfall data show similar large-scale declines from 2000 to 2012. TWS declined over most of the study area, particularly over the northern Congo. CBA also declined over 85% of the area from 2001 to 2009. Overall about 12%–28% of the forested area exhibited a significant negative trend ($P < 0.1$) for rainfall, TWS and CBA.

Figure 3 shows interannual variations of climate and vegetation variables at regional level for the period 2000–2012. EVI declined significantly by -0.021 ± 0.007 per decade ($P = 0.016$; 2000–2012) and -0.038 ± 0.009 ($P = 0.002$; 2003–2012). Rainfall, TWS, CBA and VOD also

[1]Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York (SUNY), Albany, New York 12222, USA. [2]I. M. Systems Group (IMSG), National Oceanic and Atmospheric Administration/National Environmental Satellite, Data, and Information Service/The Center for Satellite Applications and Research (NOAA/NESDIS/STAR), 5830 University Research Court, College Park, Maryland 20740, USA. [3]Department of Earth and Environment, Boston University, Boston, Massachusetts 02215, USA. [4]Laboratoire des Sciences du Climat et de l'Environnement (LSCE), CEA-CNRS-UVSQ, 91191 Gif sur Yvette Cedex, France. [5]Jet Propulsion Laboratory, Pasadena, California 91109, USA. [6]ARC Centre of Excellence for Climate Systems Science & Climate Change Research Centre, University of New South Wales, Sydney, New South Wales 2052, Australia. [7]Department of Ecology, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. [8]Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China. [9]NASA Goddard Space Flight Center, Code 619, Greenbelt, Maryland 20771, USA. [10]Department of Geography, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 29599, USA. [11]School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, Anhui 230036, China. [12]Institute for the Environment, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 29599, USA.

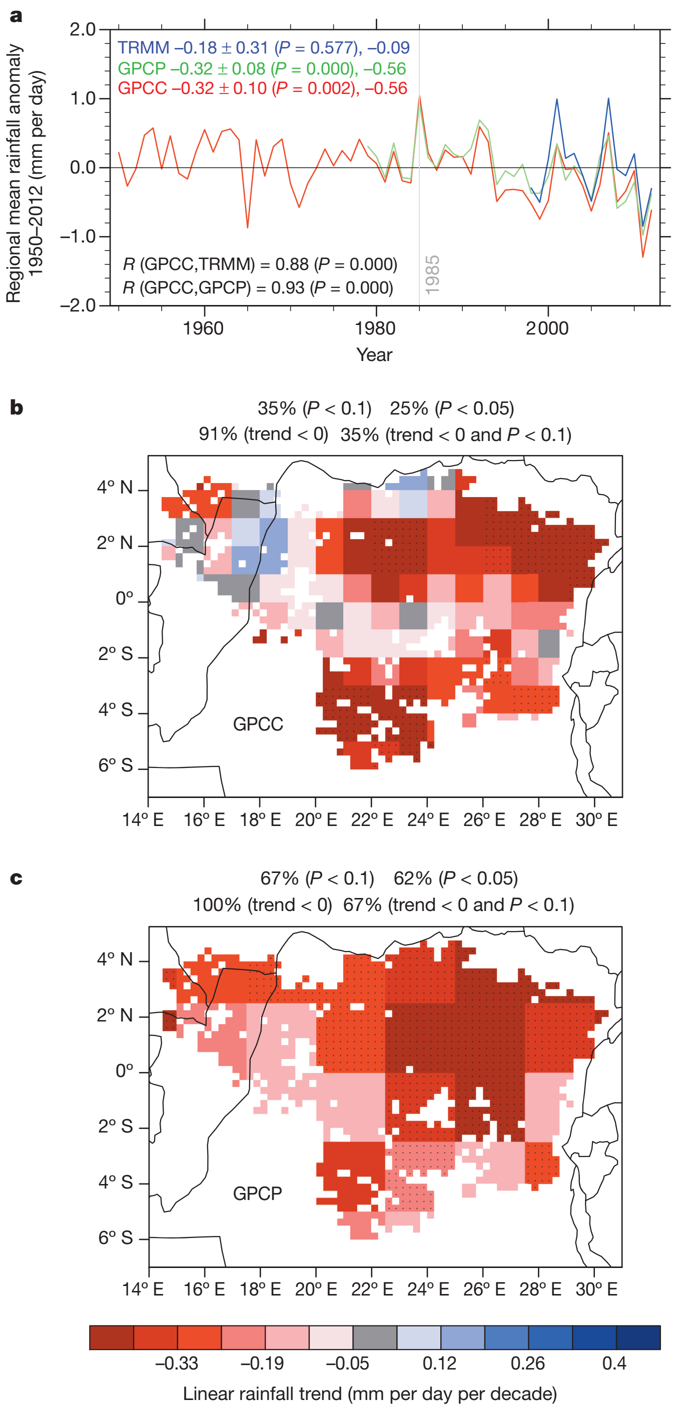


Figure 1 | April–May–June rainfall anomalies and linear trends per decade. **a**, Regional mean anomalies from three data sets (GPCC, 1950–2012; GPCP, 1979–2012; and TRMM, 1998–2012). The linear trend (±one standard deviation) and its significance level P for the period 1985–2012, the decadal difference (2003–2012 minus 1985–1994 averages) for GPCC and GPCP, and the 3-year difference (2009–2012 minus 1998–2001 averages) for TRMM are shown. The correlation coefficient R between GPCC and TRMM and between GPCC and GPCP are also shown. **b**, **c**, Spatial patterns of linear trends in GPCC (**b**) and GPCP (**c**) for the period 1985–2012 when the rainfall data showed a gradually decreasing trend. In **b** and **c**, pixels with plus symbol have a linear trend that is statistically significant at $P < 0.05$. The percentages of pixels with trends at $P < 0.05$ and $P < 0.1$ over the study region are listed, as are the percentages of pixels with negative trends.

decreased but most trends are insignificant because of strong interannual variability and short data records. However, the rainfall tended towards more negative anomalies over time and reached its lowest levels in the last few years. EVI demonstrates similar changes and so does TWS. Similar decreases are also seen for these variables in differences between the last and first three years. Most year-to-year correlations between EVI and other variables in April–May–June (no lag) are statistically significant: EVI–rainfall ($R = 0.46$–0.48, $P = 0.097$–0.114), EVI–TWS ($R = 0.68$–0.88, $P = 0.001$–0.032) and EVI–CBA ($R = 0.66$, $P = 0.037$). VOD correlates weakly with EVI ($R = 0.34$, $P = 0.312$) from 2000 to 2010, but its longer time series shows minima in the last few years (Fig. 4a). The low-frequency nature of TWS changes corresponds well to that of EVI variations and thus EVI correlates well with TWS also at the pixel scale (Extended Data Fig. 2).

The amount of radiation for plant growth and transpiration is determined by atmospheric conditions (clouds and aerosols) and latitude. COT showed a significant and widespread decreasing trend for the period 2000–2012, whereas AOT changed little, except in 2004 because of two volcanic eruptions (Extended Data Fig. 3). Consequently, PAR increased by $+7.0 \pm 3.0\%$ ($P = 0.049$) from 2003 to 2012 and this increase is stronger over the southern Congo than the northern Congo (Extended Data Fig. 4). Increasing PAR can enhance radiation-limited canopy photosynthesis[17,18] and evapotranspiration as long as the system remains energy-limited. Beyond the point where soil moisture becomes limiting, it will no longer increase transpiration and may actually enhance water stress for plant growth over the southern Congo where rainfall reduction is weaker than in the northern Congo. Consequently, LST increased over the majority of pixels during 2003–2012, particularly in the northern Congo, where strong drying trends are observed (Extended Data Fig. 4).

We also analysed the newly developed VOD data[25] for the period 1988–2010 in view of the short EVI record. The regional mean April–May–June VOD decreased steadily from 1988 to 2002, recovered slightly between 2003 and 2006, and reached the lowest levels thereafter (Fig. 4a), mirroring the low-frequency signal of rainfall variations. It decreased significantly, -0.003 ± 0.001 ($-0.3 \pm 0.1\%$) per decade ($P = 0.002$) or by -0.003 (-0.3%) between the last and first decades. The magnitude of VOD trend is small owing to its large spatial variability, and in particular, VOD represents mostly above-ground woody biomass water content[26]. At the pixel level, VOD decreased significantly ($P < 0.1$) over 38% of the study area with a larger magnitude (Fig. 4b). The annual mean VOD shows similar trends and variability (Extended Data Fig. 5). We note that VOD may be underestimated over open water areas[25]. However, the decline in rainfall would result in less standing open water spatially and temporally and thus larger increases in VOD values rather than the observed decreasing trend (Supplementary Information section A).

The lack of a strong spatial coupling between rainfall and EVI/VOD changes (Figs 1b and c, 2a–d and 4b) is not surprising given the large spatiotemporal variability of rainfall[10], the complex lagged relationships between rainfall and vegetation phenology and photosynthetic activity[1–3], and the different responses of various plant species to drought[27]. Deep roots in some forests allow trees to access ground water and delay the inception of moisture stress. Although EVI and VOD provide complementary information on vegetation dynamics, EVI is more responsive to changes in canopy cover/greenness and short-term precipitation variations whereas VOD is more sensitive to changes in woody vegetation and long-term precipitation variation[26]. Consequently there are lagged correlations between EVI/VOD in April–May–June and mean rainfall in earlier months: April–May–June EVI correlated significantly ($R = 0.69$, $P = 0.009$) with March–April–May rainfall and April–May–June VOD correlated significantly ($R = 0.49$, $P = 0.017$) with rainfall during December–January–February–March–April–May (Figs 3b and 4a). The TWS–rainfall correlations are similarly weak, because TWS in wet tropical regions represents surface and groundwater changes rather than small-scale and short-term rainfall anomalies[23,24]. In particular, the moisture available in the root zone for plant photosynthesis is determined not only by the sum of rainfall but also by runoff and evapotranspiration[24]. Over the Congo, TWS decreased most in low elevation valleys, not over the areas where rainfall decreased most (Fig. 2c-e), because local upland

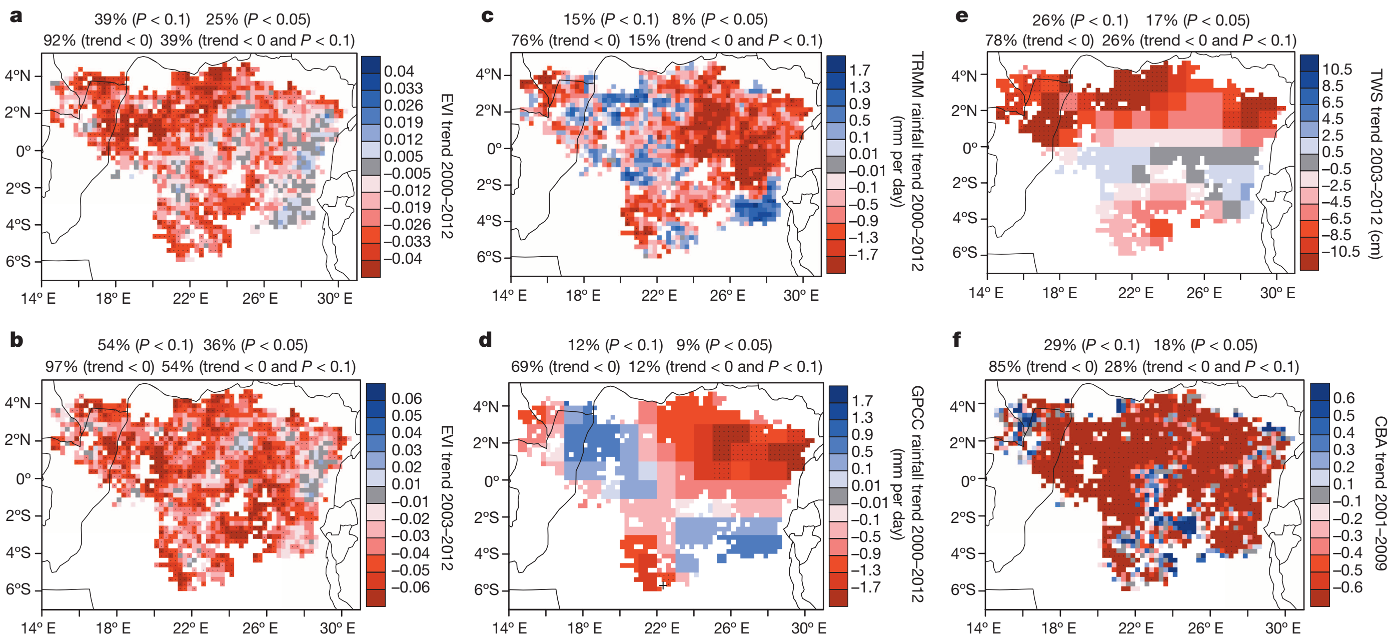


Figure 2 | Spatial patterns of linear trends per decade in April–May–June for the period 2000–2012. Pixels with plus symbols have a linear trend that is statistically significant ($P < 0.1$). The percentages of pixels with trends at $P < 0.05$ and $P < 0.1$ and the percentages of pixels with negative trends over the study region are shown. **a**, **b**, EVI. **c**, **d**, TRMM and GPCC rainfall. **e**, Ensemble mean TWS from three data processing centres. **f**, CBA. (EVI and CBA are unitless.) The CBA for 2000 was not used because the data during the period 2001–2009 in general show a decreasing trend (Fig. 3d). TWS from the three data processing centres shows similar large-scale patterns of linear trends (Extended Data Fig. 2).

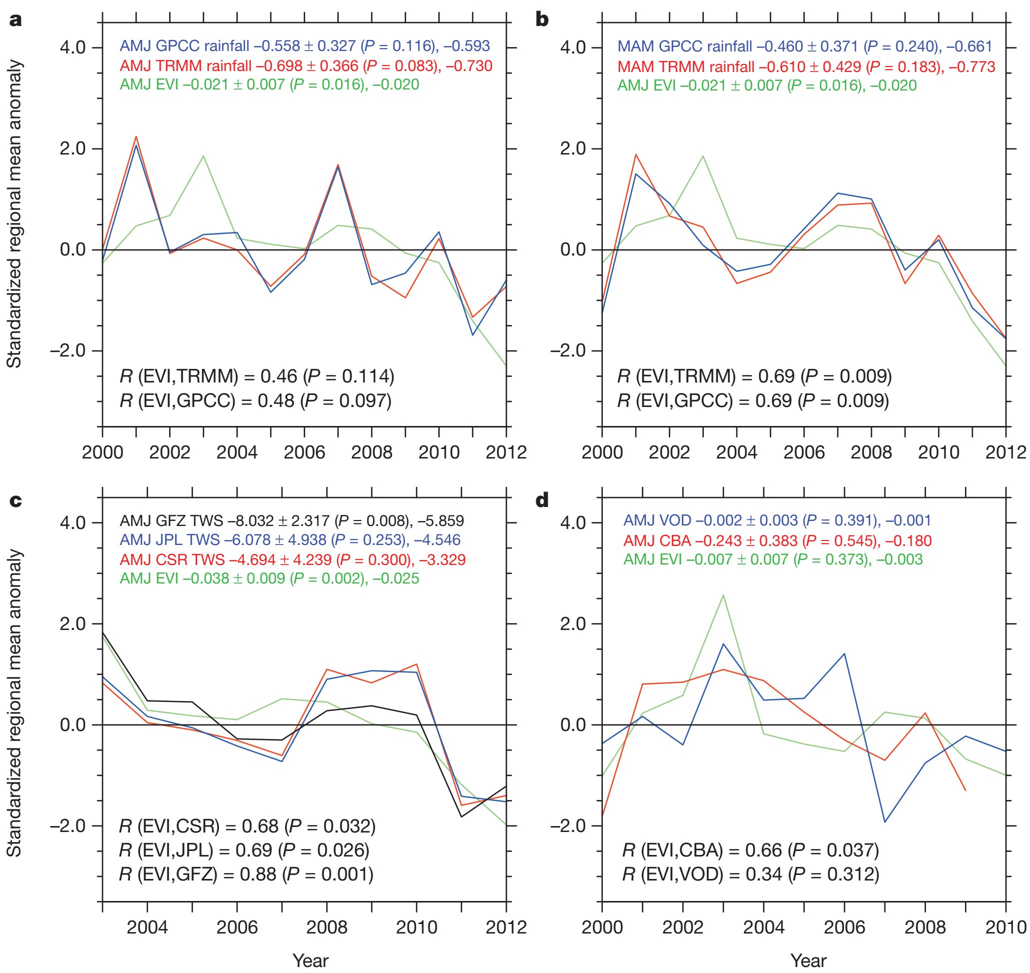


Figure 3 | Regional mean anomalies in EVI, rainfall, TWS, CBA and VOD. (VOD is unitless; rainfall has units of millimetres per day, TWS has units of centimetres.) The linear trend (±one standard deviation) per decade and its significance level P are shown, followed by the mean value difference between the last and first three years. The correlation coefficient R and its P value between EVI and other variables are shown. All variables are standardized for visualization purposes, owing to the differences in magnitude. **a**, April–May–June (AMJ) EVI and April–May–June rainfall. **b**, April–May–June EVI and March–April–May (MAM) rainfall. **c**, April–May–June EVI and April–May–June TWS from three data processing centres: the University of Texas' Center for Space Research (CSR), NASA's Jet Propulsion Laboratory (JPL) and Germany's GeoForschungsZentrum (GFZ). **d**, EVI, VOD and CBA in April–May–June.

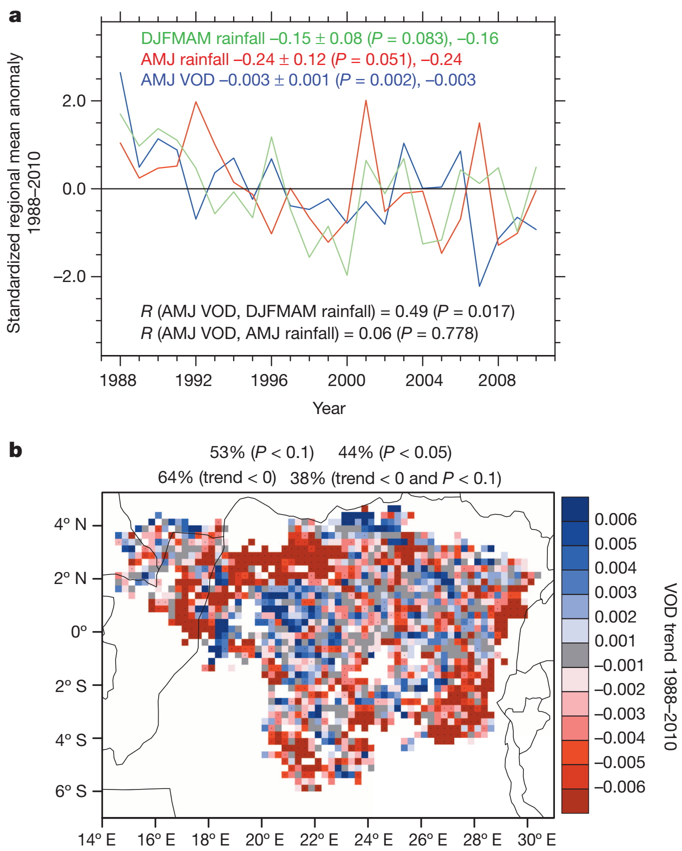


Figure 4 | VOD anomalies and linear trends per decade in April–May–June for the period 1988–2010. **a**, Regional mean VOD anomaly in April–May–June and its relationship with rainfall from GPCC in April–May–June (no lag) and DJFMAM (December–January–February–March–April–May, with lags). The linear trend (±one standard deviation), its significance level P and the decadal difference (2001–2010 minus 1988–1997 averages) are shown. The correlation coefficient R and its P value is also shown. **b**, Spatial patterns of linear trend per decade of VOD for the period 1988–2010. Pixels with plus symbols have a linear trend that is statistically significant at $P < 0.05$; the percentages of pixels with trends at $P < 0.05$ and $P < 0.1$ and the percentages of pixels with negative trends over the study region are shown.

runoff is the main source of the Congo wetland water[28]. Furthermore, the effects of long-term drought on vegetation are more complex than short-term drought[11] and different satellite products measure different properties of vegetation and moisture[11,25] at different spatiotemporal resolutions, which make it difficult to infer a strong coupling between vegetation and moisture parameters at pixel level. Nevertheless, the combined areas with a decreasing trend in rainfall and TWS indicate a widespread and increasing water deficit, which agrees consistently with the large-scale EVI decreases.

Most of the EVI decline reflects real changes in vegetation characteristics rather than data errors[7], residual atmospheric artefacts due to aerosol and cloud contaminations[6], sun-view angle effects, sensor degradation, deforestation, and natural fires (Extended Data Figs 6–9 and Supplementary Information sections B–E). The persistent browning of the Congolese forests might reflect a slow adjustment to the long-term drying trend, rather than a response to episodic events such as the Amazon droughts. It is generally consistent with the gradual temporal changes in moisture, vegetation and radiation parameters observed from several independent satellite products. In particular, VOD and CBA, together with high-resolution satellite images from Landsat (Extended Data Fig. 10 and Supplementary Information section D), only show small and gradual changes in canopy structure instead of large-scale tree mortality as in the Amazon[8]. The large decline in EVI and VOD in recent years is also paralleled by an acceleration of climate drivers such as PAR, rainfall and TWS changes. Furthermore, the persistent decrease in COT and strong increase in PAR support the observed decrease in rainfall, warming of the forested landscape and the increase in evapotranspiration demand over the Congo.

Water deficit stress will be manifested in trees as less turgid leaves, and if the stress is beyond some threshold, the trees may drop their leaves. This behaviour can be detected as a decline in EVI. A slow decline of EVI over time implies loss of photosynthetic capacity. Such a temporary loss due to short-term drought may have minimal effects on vegetation because the trees may leaf out as soon as rainfall recovers. However, under the stress of an increased severity of water deficit, the gradual loss of photosynthetic capacity and water content over a long period may alter forest species composition and structure[1,2,14] and thus affect biodiversity and carbon storage[1,29]. Drier conditions may favour deciduous trees at the expense of evergreen trees[2,14]. For example, over a 20-year period of exposure to a drying trend, the Ghanaian forest species shifted from wetter-forest-affiliated vegetation to deciduous and drier-forest canopy species[14]. Detection and attribution of such impacts require long-term ground observations and drought manipulative experiments (as for the Amazon forests[30]); these, however, are not available for Congolese forests, and should be a research priority.

METHODS SUMMARY

We analyse gridded data from ground observations and different satellite sensors over the Congo (see Methods for details): (1) MODIS products of EVI[15] (monthly, 2000–2012), LST (monthly, 2003–2012), and land cover (yearly, 2001–2012) at 0.05° resolution, and AOT and COT at 1° resolution (monthly, 2000–2012); (2) monthly rainfall from GPCC[20] (1°, 1950–2012), GPCP[21] (2.5°, 1979–2012) and TRMM[22] (0.25°, 1998–2012); (3) monthly TWS[23] and PAR (1°, 2003–2012); and (4) monthly data of VOD[25] (0.25°, 1988–2010) and CBA[11] (0.04°, 2000–2009). We focused only on the intact Congolese forest pixels (5° N–6° S, 14° E–31° E) during April–May–June using a high-quality EVI mask based on EVI quality assurance information. For each year, monthly high-quality 0.05° EVI anomalies were aggregated into one single seasonal (April–May–June) image at 0.25° resolution. The other variables were similarly averaged into seasonal anomalies and then re-projected into the study region at 0.25° resolution. Regional and spatial aggregations were applied only to the masked high-quality EVI pixels. Two methods were used to quantify the temporal changes at both the pixel level and the regional level: first, a mean value difference between two individual periods; and second, a linear trend estimated using ordinary least squares over a certain time period. For the first method, the two periods are defined as the last and first three years if the data record is less than 20 years long, and the last and first decades otherwise. For the second method, the Student's t-statistic was calculated to quantify the probability P of whether the trend is statistically significantly different from zero or simply due to random noise. Both methods should provide consistent results if there is a persistent trend in the data. A linear correlation and its significance level (P value) were calculated between two time series to quantify their association. The acronyms and abbreviations are listed in Supplementary Information Supplementary Table 4.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August 2013; accepted 21 March 2014.
Published online 23 April 2014.

1. Lewis, S. L. Tropical forests and the changing earth system. *Phil. Trans. R. Soc. Lond. B* **361,** 195–210 (2006).
2. Enquist, B. J. & Enquist, C. A. F. Long-term change within a neotropical forest: assessing differential functional and floristic responses to disturbance and drought. *Glob. Change Biol.* **17,** 1408–1424 (2011).
3. Lewis, S. L. *et al.* Above-ground biomass and structure of 260 African tropical forests. *Phil. Trans. R. Soc. Lond. B* **368,** 20120295 (2013).
4. Brando, P. M. *et al.* Seasonal and interannual variability of climate and vegetation indices across the Amazon. *Proc. Natl Acad. Sci. USA* **107,** 14685–14690 (2010).
5. Xu, L. *et al.* Widespread decline in greenness of Amazonian vegetation due to the 2010 drought. *Geophys. Res. Lett.* **38,** L07402 (2011).
6. Samanta, A., Ganguly, S., Vermote, E., Nemani, R. R. & Myneni, R. B. Why is remote sensing of Amazon forest greenness so challenging? *Earth Interact.* **16,** 1–14 (2012).
7. Atkinson, P. M., Dash, J. & Jeganathan, C. Amazon vegetation greenness as measured by satellite sensors over the last decade. *Geophys. Res. Lett.* **38,** L19105 (2011).

8. Saatchi, S. *et al.* Persistent effects of a severe drought on Amazonian forest canopy. *Proc. Natl Acad. Sci. USA* **110,** 565–570 (2013).
9. Morton, D. C. *et al.* Amazon forests maintain consistent canopy structure and greenness during the dry season. *Nature* **506,** 221–224 (2014).
10. Malhi, Y. & Wright, J. Spatial patterns and recent trends in the climate of tropical rainforest regions. *Phil. Trans. R. Soc. Lond. B* **359,** 311–329 (2004).
11. Asefi-Najafabady, S. & Saatchi, S. Response of African humid tropical forests to recent rainfall anomalies. *Phil. Trans. R Soc. B* **368,** 20120306 (2013).
12. Adams, J. The distribution and variety of equatorial rain forest. http://www.esd.ornl.gov/projects/qen/rainfo.html (1998).
13. Ashton, M. S., Tyrrell, M. L., Spalding, D. & Gentry, B. (eds) *Managing Forest Carbon in a Changing Climate* (Springer, 2012).
14. Fauset, S. *et al.* Drought-induced shifts in the floristic and functional composition of tropical forests in Ghana. *Ecol. Lett.* **15,** 1120–1129 (2012).
15. Huete, A. *et al.* Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **83,** 195–213 (2002).
16. Myneni, R. B., Hall, F. G., Sellers, P. J. & Marshak, A. L. The meaning of spectral vegetation indices. *IEEE Trans. Geosci. Remote Sens.* **33,** 481–486 (1995).
17. Huete, A. R. *et al.* Amazon rainforests green-up with sunlight in dry season. *Geophys. Res. Lett.* **33,** L06405 (2006).
18. Nemani, R. R. *et al.* Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300,** 1560–1563 (2003).
19. de Wasseige, C., Bastin, D. & Defourny, P. Seasonal variation of tropical forest LAI based on field measurements in Central African Republic. *Agric. For. Meteorol.* **119,** 181–194 (2003).
20. Schneider, U. *et al.* GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* **115,** 15–40 (2014).
21. Adler, R. F. *et al.* The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeorol.* **4,** 1147–1167 (2003).
22. Huffman, G. J. *et al.* The TRMM Multi-satellite Precipitation Analysis: quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol.* **8,** 38–55 (2007).
23. Crowley, J. W., Mitrovica, J. X., Bailey, R. C., Tamisiea, M. E. & Davis, J. L. Land water storage within the Congo Basin inferred from GRACE satellite gravity data. *Geophys. Res. Lett.* **33,** L19402 (2006).
24. Reager, J. T. & Famiglietti, J. S. Characteristic mega-basin water storage behavior using GRACE. *Wat. Resour. Res.* **49,** (2013).
25. Liu, Y. Y., van Dijk, A. I. J. M., McCabe, M. F., Evans, J. P. & de Jeu, R. A. M. Global vegetation biomass change (1988–2008) and attribution to environmental and human drivers. *Glob. Ecol. Biogeogr.* **22,** 692–705 (2013).
26. Andela, N., Liu, Y. Y., van Dijk, A. I. J. M., de Jeu, R. A. M. & McVicar, T. R. Global changes in dryland vegetation dynamics (1988–2008) assessed by satellite remote sensing: comparing a new passive microwave vegetation density record with reflective greenness data. *Biogeosciences* **10,** 6657–6676 (2013).
27. Lucht, W., Schaphoff, S., Erbrecht, T., Heyder, U. & Cramer, W. Terrestrial vegetation redistribution and carbon balance under climate change. *Carbon Balance Manag.* **1,** 6 (2006).
28. Lee, H. *et al.* Characterization of terrestrial water dynamics in the Congo Basin using GRACE and satellite radar altimetry. *Remote Sens. Environ.* **115,** 3530–3538 (2011).
29. Valentini, R. *et al.* The full greenhouse gases budget of Africa: synthesis, uncertainties and vulnerabilities. *Biogeosci. Discuss.* **10,** 8343–8413 (2013).
30. Nepstad, D. C., Tohver, I. M., Ray, D., Moutinho, P. & Cardinot, G. Mortality of large trees and lianas following experimental drought in an amazon forest. *Ecology* **88,** 2259–2269 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the NOAA NESDIS project (NA11NES4400010) and by the startup funds provided by the University at Albany, State University of New York. R.B.M. was funded by NASA's Earth Science Division. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official NOAA or US Government position, policy, or decision. H.C. is supported by the National Natural Science Foundation of China (grant number 41230422).

Author Contributions L.Z. and Y.T. contributed the central idea, analysed most of the data, and wrote the initial draft of the paper. The remaining authors contributed to refining the ideas, carrying out additional analyses and finalizing this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.Z. (lzhou@albany.edu).

METHODS

Data. This research analyses a variety of climate and vegetation data from ground observations and several independent satellite sensors (optical, thermal, passive and active microwave and gravity) over the Congo basin. We use (1) rainfall, terrestrial water storage (TWS), aerosol optical thickness (AOT), cloud optical thickness (COT), photosynthetically active radiation (PAR) and land surface temperature (LST) as climate drivers; and (2) enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), vegetation optical depth (VOD) and canopy backscatter anomaly (CBA) as vegetation variables. Unlike optical and thermal remote sensing products (for example, NDVI, EVI and LST), microwave retrievals (for example, VOD and CBA) are not affected by atmospheric cloud cover and aerosol conditions. The low energy of microwave emission does, however, require a larger footprint, resulting in a relatively coarser spatial resolution of the microwave data than the optical and thermal data used in this study. The gravity sensors from the Gravity Recovery and Climate Experiment (GRACE) map the Earth's gravity fields by making accurate measurements of the distance between the two satellites and thus provide information about the distribution and flow of water mass within the Earth and its surroundings.

MODIS data. We use the globally validated Collection 5 MODerate resolution Imaging Spectroradiometer (MODIS) products of EVI[15] and NDVI[15] (monthly, 2000–2012), LST[31] (monthly, 2003–2012) and land cover and percentage vegetation cover[32,33] (yearly, 2001–2012) at 0.05° spatial resolution, and AOT[34,35] and COT[34,36] at 1° resolution (monthly, 2000–2012). The NDVI and EVI (MOD13C2), LST (MYD11C3), and land cover (MCD12C1) data were obtained from https://lpdaac.usgs.gov/get_data. The AOT and COT data (MOD08M3) were obtained from http://ladsweb.nascom.nasa.gov/data/search.html.

The MODIS EVI data (MOD13C2) are used primarily because EVI is more sensitive to dense forests than other vegetation indices such as NDVI and has been used widely in recent studies on tropical rainforests[4–7]. The MODIS EVI algorithm[15] uses the MODIS surface reflectance data in the blue, red and near-infrared spectral bands (which are corrected for molecular scattering, ozone absorption and aerosols) as its input. The blue band removes residual atmospheric contamination caused by smoke and sub-pixel thin clouds. A feedback adjustment is used to minimize canopy background variations and to enhance its sensitivity from sparse vegetation to very dense forests[15,17]. When properly filtered to remove atmospheric aerosol and cloud effects, the MODIS EVI data do not saturate, even over dense forests, and correlate well with leaf area, leaf biomass, canopy chlorophyll content, canopy photosynthetic activity and primary productivity[16–18]. The gridded EVI data sets include pixel-level quality assurance flags together with statistics of the EVI quality and input data. MOD13C2 represents the best retrievals possible from cloud-free high-quality spatial composites of the gridded 16-day 1-km EVI product (MOD13A2) at local solar time ~10:30 for the period 2000–2012 from the Terra satellite. It incorporates a quality assurance filter scheme that removes lower-quality and cloud-contaminated pixels during spatial aggregation. The accuracy of the MODIS EVI has been assessed over a widely distributed set of locations and time periods via validation efforts, with an error of ±0.01 (ref. 7). The MODIS NDVI, which was calculated and processed similarly to the EVI by the same MODIS algorithm[15], is used to assess the sun-view angle effects (see Supplementary Information).

The MODIS daytime LST[31] (MYD11C3) at local solar time ~13:30 for the period 2003–2012 from the Aqua satellite is used to quantify LST variations related to changes in vegetation and other land surface properties. Satellite-derived LST measures the temperature of the Earth's surface thermal emission. It is closely related to vegetation density/type and hence has been used to monitor deforestation and land cover change[37,38]. The MODIS LST has four measurements from the Terra and Aqua satellites. Here we choose the daytime Aqua observations because the LST contrast between vegetation and non-vegetation is expected to be strongest at ~13:30 than the other three MODIS measurements. However, it may have larger uncertainties than the other three measurements because clouds are more often seen over land during the afternoon than in the morning.

The MODIS land cover map[32,33] (MCD12C1) is available from 2001 and 2012 and can be used to identify and quantify large-scale changes in vegetation type and percentage vegetation cover. It provides a suite of land cover types by mapping global land cover using spectral and temporal information derived from MODIS using a supervised decision-tree classification method based on a database of high-quality land cover training sites[32].

The MODIS level 3 joint aerosol/water-vapour/cloud products[34–36] (MOD08M3) provide AOT and COT data collected from the Terra satellite for the period 2000–2012. We used the fields of *Optical_Depth_Land_And_Ocean_Mean_Mean*, which contains AOT at 0.55 μm for both ocean (best) and land (corrected), and *Cloud_Optical_Thickness_Combined_Mean_Mean*, which contains the monthly mean values of daily mean of combined COT (liquid plus ice), from the MOD08M3 product.

Rainfall data. We use three gridded monthly rainfall data sets from the Global Precipitation Climatology Centre (GPCC)[20] at 1° resolution (1950–2012), the Global Precipitation Climatology Project (GPCP)[21] at 2.5° resolution (1979–2012) and the Tropical Rainfall Measuring Mission (TRMM) and other satellites[22] at 0.25° resolution (1998–2012). The GPCC data provides the GPCC's most accurate *in situ* land surface precipitation analysis product by combining the Full Data V6 Product (1901–2010), based on quality-controlled data from 67,200 stations worldwide that feature record durations of ten years or longer, with the V4 monitoring product (2011–2012), based on quality-controlled data from 7,000 stations. The GPCP version 2.2 data provides combined gauge measurements and satellite-derived precipitation by taking advantages of the strengths of each of several sources into a final merged product. The 3B43 version 7 TRMM data provides the best-estimate precipitation rate and root-mean-square precipitation-error estimates by combining four independent precipitation fields. Despite their differences in data source, duration, spatial resolution and processing, these three rainfall data sets show similar and consistent rainfall variability during April–May–June over the study region (Fig. 1). GPCC and TRMM have higher spatial resolution and thus are mostly used in our analysis. We use the former (1950–2012) to examine the long-term variability of rainfall and the latter (2000–2012) to examine spatial patterns of rainfall in the last 13 years. The data sets were obtained from http://www.esrl.noaa.gov/psd/, http://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html and http://mirador.gsfc.nasa.gov/collections/TRMM_3B43__007.shtml, respectively.

We had previously planned to include the gridded Climatic Research Unit (CRU) rainfall data[39] (CRU_TS_3.21) for the period 1950–2012 in our analysis as well. However, we found that the data of CRU differed from those of GPCC, GPCP and TRMM in the last decade over our study region owing to the insufficient ground observations used in the CRU. Asefi-Najafabady and Saatchi[11] also compared the rainfall quality of CRU and TRMM and found large uncertainties in CRU over the African forests for the period after 1990 owing to paucity of station data, whereas the TRMM products have relatively low bias in magnitude and errors in capturing rainfall spatial patterns.

TWS data. We use the gridded monthly Release 05 (RL05) TWS data from GRACE[23,40,41] at 1° resolution (2003–2012). It has three products, processed by the University of Texas' Center for Space Research, NASA's Jet Propulsion Laboratory and Germany's GeoForschungsZentrum, and was obtained from http://grace.jpl.nasa.gov/data/gracemonthlymassgridsland/. TWS represents the total water storage changes in wetlands, rivers, ground water, soil moisture and wet biomass in terms of anomalies relative to a mean total storage value. It is determined by the balance between source (precipitation) and sink (runoff, evapotranspiration) terms, and so a change in either source or sink will change TWS[23,24]. Although two areas may receive similar amounts of precipitation, the resulting storage response may differ owing to differences in temperature and the characteristics of land surface properties (for example, vegetation type/amount, soil depth and porosity, and topography) with which the water interacts[24]. Given that a 200-km-wide Gaussian filter was applied to the data, TWS quantifies a large-scale and low-frequency signal of TWS anomalies. The year-to-year TWS values from the University of Texas' Center for Space Research and NASA's Jet Propulsion Laboratory are similar but differ slightly from those from Germany's GeoForschungsZentrum (Fig. 3c and Extended Data Fig. 2). Our understanding and validation of TWS products are insufficient for a precise assessment of the quality among these three data sets, so here we simply use the ensemble-mean TWS with equal weights to illustrate the spatial pattern of TWS trends (Fig. 2e). Nevertheless, the results from the three data sets generally show similar decreasing trends in spatial patterns (Fig. 2e and Extended Data Fig. 2) and interannual variations at the regional level (Fig. 3b).

Radiation data. We use gridded monthly data of downward PAR at the surface from Clouds and Earth's Radiant Energy System (CERES)[42] at 1° resolution (2003–2012). The CERES products contain monthly mean direct and diffuse surface fluxes (CERES_SYN1deg_Ed3A) of PAR under all-sky and clear-sky conditions, which were obtained from http://ceres.larc.nasa.gov/. The total incident all-sky PAR was calculated by summarizing direct and diffuse PAR fluxes. The total incident shortwave radiation shows changes similar to PAR's but with a larger magnitude (not shown for brevity).

VOD data. We use gridded monthly VOD data retrieved from the Special Sensor Microwave Imager (SSM/I) and the Advanced Microwave Scanning Radiometer—Earth Observing System (AMSR-E)[25] at 0.25° resolution. VOD is a radiometric parameter retrieved from passive microwave satellite observations which can penetrate deeper into the canopy than can optical remotely sensed data such as EVI and is directly proportional to total vegetation water content in all the aboveground biomass, including both woody and leafy components[43–45]. Unlike EVI, VOD is sensitive to water rather than chlorophyll and hence to both photosynthetic and non-photosynthetic aboveground biomass[46,47]. The VOD retrievals from the SSM/I (January 1988–June 2002) and AMSR-E (July 2002–December 2010) sensors[25] were merged to create a single continuous long-term time series by using a cumulative distribution frequency matching technique[48–50]. Because of its higher retrieval accuracy, the AMSR-E VOD was selected as the reference with which to adjust the SSM/I

data. We note that the merging process has little influence on the long-term trends, that is, the trends from their original sources are kept unchanged. A comprehensive study has demonstrated that this long-term VOD data set is able to monitor global changes in total aboveground vegetation water content and biomass over various land cover types from grasslands to tropical forests[25]. The AMSR-E sensor stopped working in October 2011, which limits our analysis up to 2010.

CBA data. We also use gridded CBA of the SeaWinds scatterometer data onboard Quick Scatterometer (QSCAT)[11] (2000–2009) at 0.04° resolution. CBA provides a reliable remote sensing technique to monitor the impact of climate on tropical forests, especially in detecting large-scale tree mortality and tree leaf abscission. QSCAT is an active radar sensor operating at microwave frequency (13.4 GHz) and provides daily (6.00 and 18.00) measurements of the backscatter signal from forest canopies. QSCAT backscatter measurements over dense tropical forests are sensitive only to the top canopy structure and moisture and contain little information about the underlying soil moisture. The global wall-to-wall acquisition of QSCAT data stopped in November 2009, limiting our analysis up to 2009.

Data processing. The African rainforests span the equatorial region by nearly seven degrees from north to south, but some forested regions such as in West Africa have extensive cloud and aerosol contaminations, as can be inferred from optical remote products such as MODIS NDVI and EVI. We chose our study region and study period after carefully assessing the EVI data quality (Extended Data Fig. 1). First, the MODIS land cover classification and percentage forest cover data were used to define intact forest canopies in the Congo basin. Only forested pixels at 0.05° resolution that have a forest cover exceeding 80% and have no land cover/use change detected during the period 2001–2012 are considered. Second, the quality assurance flags of each monthly EVI composite were checked to count the total number of high-quality EVI composites on every pixel at 0.05° resolution for the 13-year period. A high-quality EVI composite was defined as one without the presence of clouds (adjacent clouds, mixed clouds and shadows) and aerosol loadings that typically corrupt EVI (climatology and high aerosols)[6]. The three-month period of April–May–June shows the highest percentage of high-quality EVI pixels (or the least contamination of aerosols and clouds) over the 13-year period and also coincide with the first peak season of EVI and rainfall, so we focus our study period on April–May–June only. Third, the quality assurance statistics were used to create a high-quality EVI mask at 0.05° resolution. The masked pixels contain only those pixels having at least 80% of monthly composites in April–May–June (or at least 31 of the total 39 monthly composites) for the 13-year study period, which represent at least ten years of high-quality data. Only the masked high-quality pixels are considered so that the same group of EVI pixels is analysed for the entire study period. Fourth, for each of the masked pixels, a monthly high-quality EVI anomaly time series in April–May–June was first created by removing the monthly high-quality EVI climatology for the period 2000–2012 and then aggregated into one single April–May–June high-quality EVI anomaly image for each year. Fifth, the 0.05° April–May–June high-quality anomaly images were spatially aggregated into coarser resolution images at 0.25° to enhance the pixel-level signal-to-noise ratio and also match the spatial resolution of TRMM rainfall. We only consider those 0.25° resolution pixels that (1) have at least five sub-pixels that are masked as high-quality at 0.05° resolution (to ensure adequate samples for spatial averaging) and (2) are defined as intact forested pixels in the first step. Some isolated pixels near the border of enclosed forests were removed from the 0.25° mask map to minimize human impacts. The above steps lead to the choice of our study region over only the intact forest pixels with high-quality EVIs (5° N–6° S, 14° E–31° E) over the Congo basin, which in total consists of 1,438 pixels at 0.25° resolution (Extended Data Fig. 1).

Like the MODIS EVI, the data of rainfall, NDVI, LST, TWS, AOT, COT, PAR, VOD and CBA at different spatial resolutions were temporally averaged into anomalies in April–May–June and other seasons and then re-projected into our study region at 0.25° resolution. For the data with a spatial resolution coarser than 0.25°, nearest-neighbour interpolation was used. Regional and spatial aggregations were applied only over the masked 0.25° pixels where the MODIS data has the most high-quality EVI composites.

The VOD data can provide complementary information on vegetation dynamics when compared to other satellite products[25,26] and especially it can be used to monitor global vegetation biomass change over various land cover types[25]. However, as a new product, the VOD has not been validated against ground observations. For example, we found that there are few strong VOD anomalies with a value exceeding 0.2 or less than −0.2 (or 20% of the absolute regional mean VOD value) at local scales, which represent 0.04% of all the data, possibly owing to uncertainties and errors from measurements and retrievals or local fires. Instead of excluding these values, which would create missing data, we treat them as local-scale noise. We use empirical orthogonal function (EOF) analysis to isolate the VOD variance associated with the smaller scales. The EOF method has been extensively used to analyse the spatial and temporal variability of geophysical fields by decomposing the data into a set of spatial patterns of variability (referred to as EOF modes) and corresponding time variations (referred to as EOF time series)[51,52]. Its goal is to express the signal in terms of a relatively small number of EOFs to describe as much of the original information as possible. The EOF modes show the spatial structure of the major factors that can account for the temporal variations, which represent spatial variability. The EOF time series tells us how the amplitude of each EOF mode varies with time. The first few EOFs contain the majority of data variance and the highest spatial coherence and the last few EOFs are noise-dominated and have the least variance. Therefore, the inversion of the EOF transform using only the first few EOFs provides a noise-filtered data set. We used the first 12 EOFs to restructure the VOD data, which explains 90% of the data variance. The restructured VOD looked almost identical to the original data, except for the smoothing of few strong VOD anomalies.

Quantification of temporal changes. When choosing methods with which to quantify the temporal changes in EVI and other variables we should consider how rainfall changes with time in our study region. Unlike short-term and intense droughts that have recently occurred in the Amazon forests[4–8], the African forests have experienced a long-term and gradual rainfall reduction[10,11]. The effects of such rainfall changes on vegetation are more complex than short-term droughts because forest composition and structure may change over time[1,2,14]. So we expect to see small and gradual changes in moisture properties (and, hence, vegetation), instead of large-scale tree mortality similar to what has occurred in the Amazon[8].

Here we use two simple methods to quantify the gradual temporal change in a variable $y(t)$, which represents EVI or any other variables over time t. The first method is to calculate the differences in the mean values of $y(t)$ between two individual periods. The two periods are defined as the last and first three years if the data record is less than 20 years. Otherwise, we use the last and first decades. The first method is straightforward and effective if $y(t)$ contains a deterministic trend, but the resulting change depends on the chosen time periods if $y(t)$ has strong interannual variations. The second method is to estimate the linear time trend of $y(t)$ using ordinary least squares over a certain time period (that is, the estimated change per decade). Although most of the remotely sensed data are only available for 10–13 years, the second method can be a good indicator if $y(t)$ shows a persistent trend (either upward or downward) with time. In particular, it can provide the P value of the Student's t statistic that quantifies the probability of whether the estimated trend is statistically significantly different from zero or simply due to random noise. Nevertheless, both methods should provide consistent results if there is a persistent trend in the data as shown in most of the variables in our analysis.

Both methods were used for time series analysis in our study, but the first method was not used in the spatial pattern analysis because the pixel-level time series is noisier than the regional mean data. So we express the temporal change mostly as a linear trend—which is one simple way to quantify the gradual change while reducing year-to-year data noise. We do realize the limitation of estimating a linear trend for a short time series of 10–13 years for most of the satellite data. Hence, the estimated trend only applies to the study period and thus should not be extrapolated linearly over longer periods.

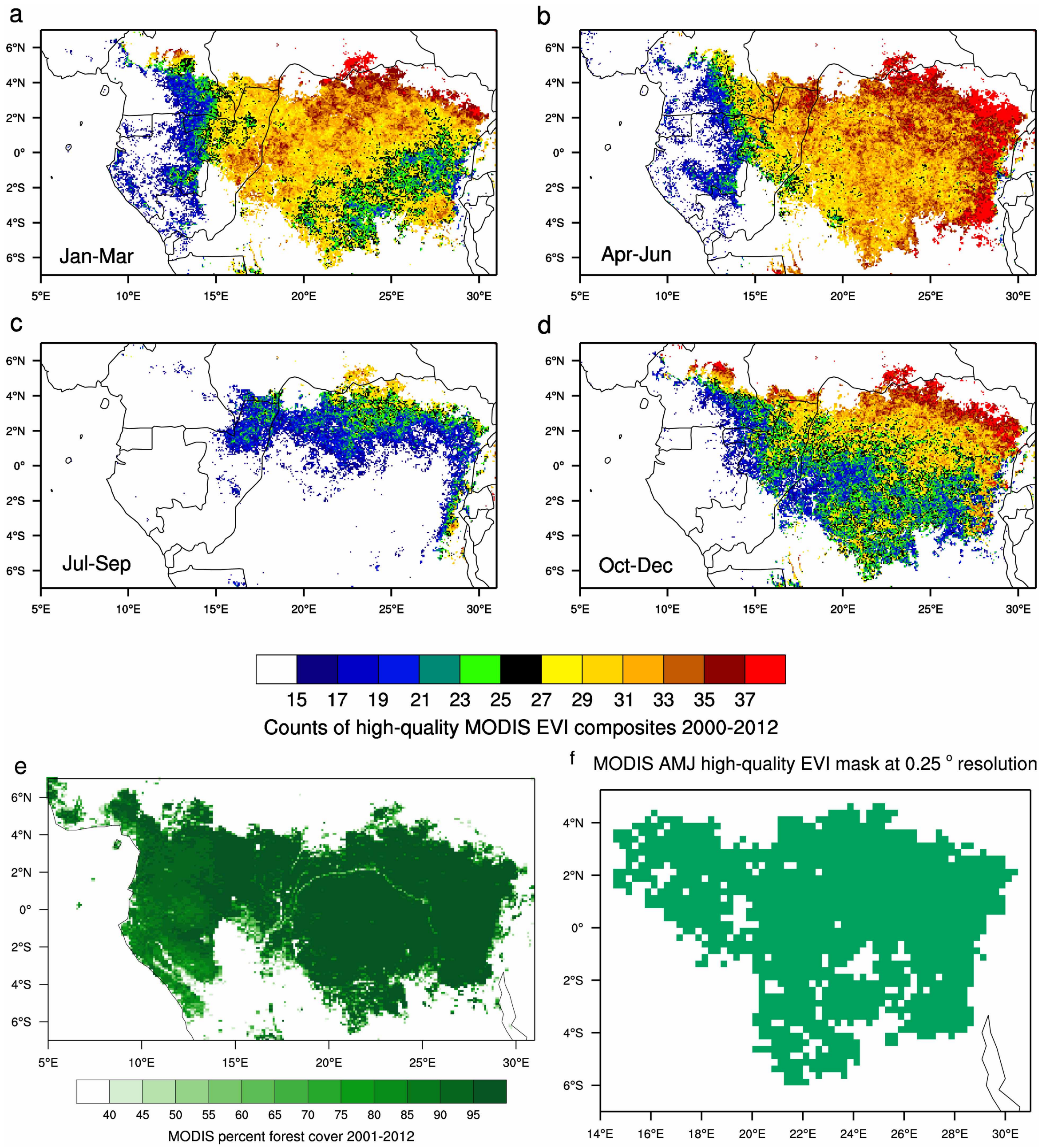
There is another method that has been used extensively to detect and quantify the drought impacts in studies of the Amazon rainforests and has proved to be very effective[5–8]. This method calculates the standardized anomaly for one particular drought year $y(t_0)$ as the departure of $y(t)$ from the climatological mean of $y(t)$ and normalized by the standard deviation of $y(t)$, while the measurement from the particular drought year $y(t_0)$ is excluded from the calculations of mean and standard deviation. It requires us to identify abnormal years versus normal years, which is relatively easy in the Amazon given the two historic drought events that occurred within the last 13 years. However, our study region has not experienced significant widespread droughts similar to what occurred in the Amazon[8], and the drought-affected areas are smaller in spatial extent than the Amazon and vary by periods[11]. Also, the decreasing trend in EVI and other variables often makes it difficult to define one particular year as an abnormal year because the last few years always have lower values than the first few years. Consequently, this method is not used in our analysis.

Correlation analysis. The linear correlation coefficient (Pearson's R) was calculated between two time series to quantify their concurrent and lagged association. The significance level of correlation (P value) is estimated using a two-tailed Student's t-test. The null hypothesis is that the two variables used to calculate R are independent (that is, $R = 0.0$).

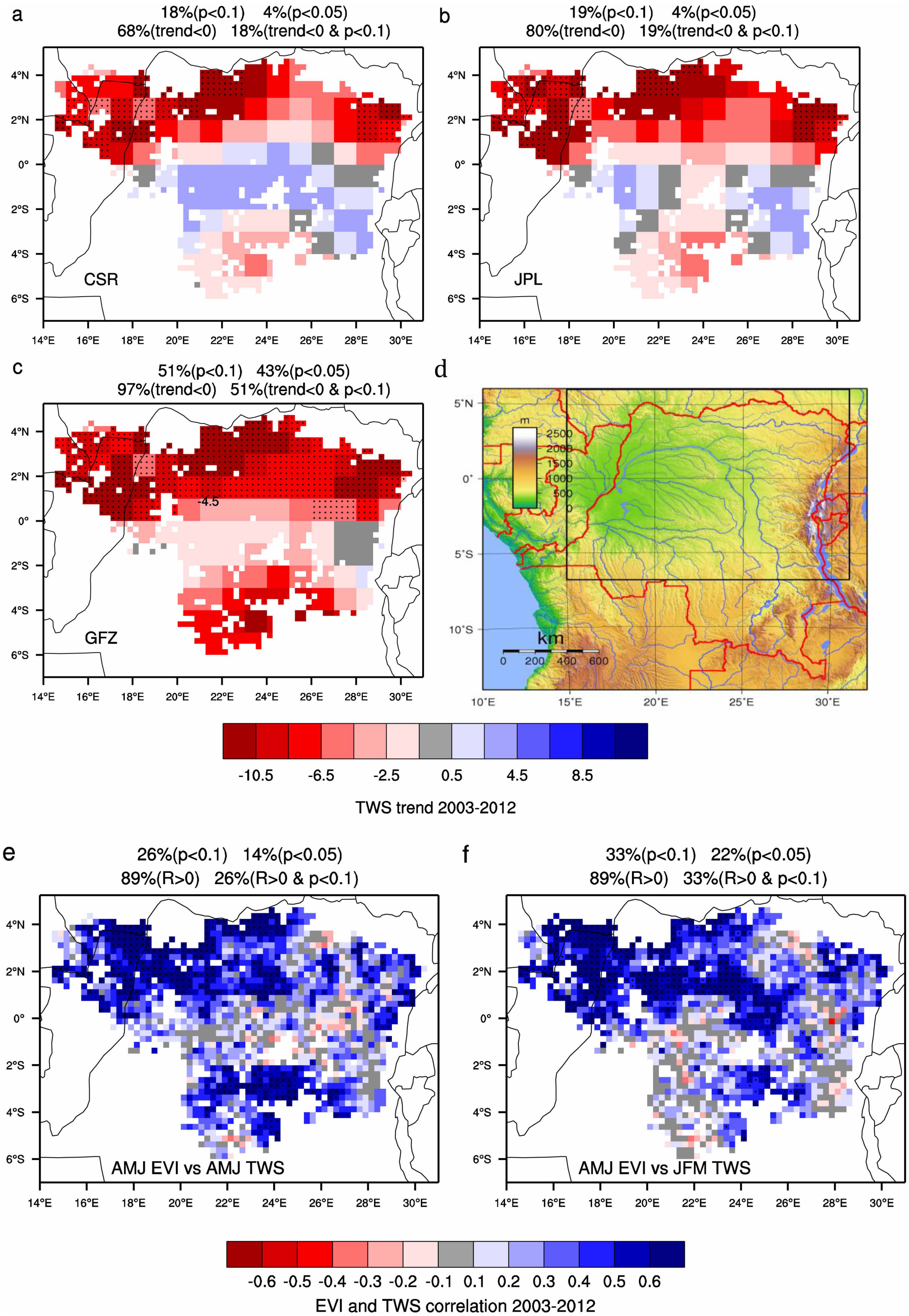
The acronyms and abbreviations are listed in Supplementary Table 4.

31. Wan, Z. New refinements and validation of the MODIS land surface temperature/emissivity products. *Remote Sens. Environ.* **112,** 59–74 (2008).
32. Friedl, M. A. *et al.* MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **114,** 168–182 (2010).
33. DiMiceli, C. M. *et al. Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000-2010,*

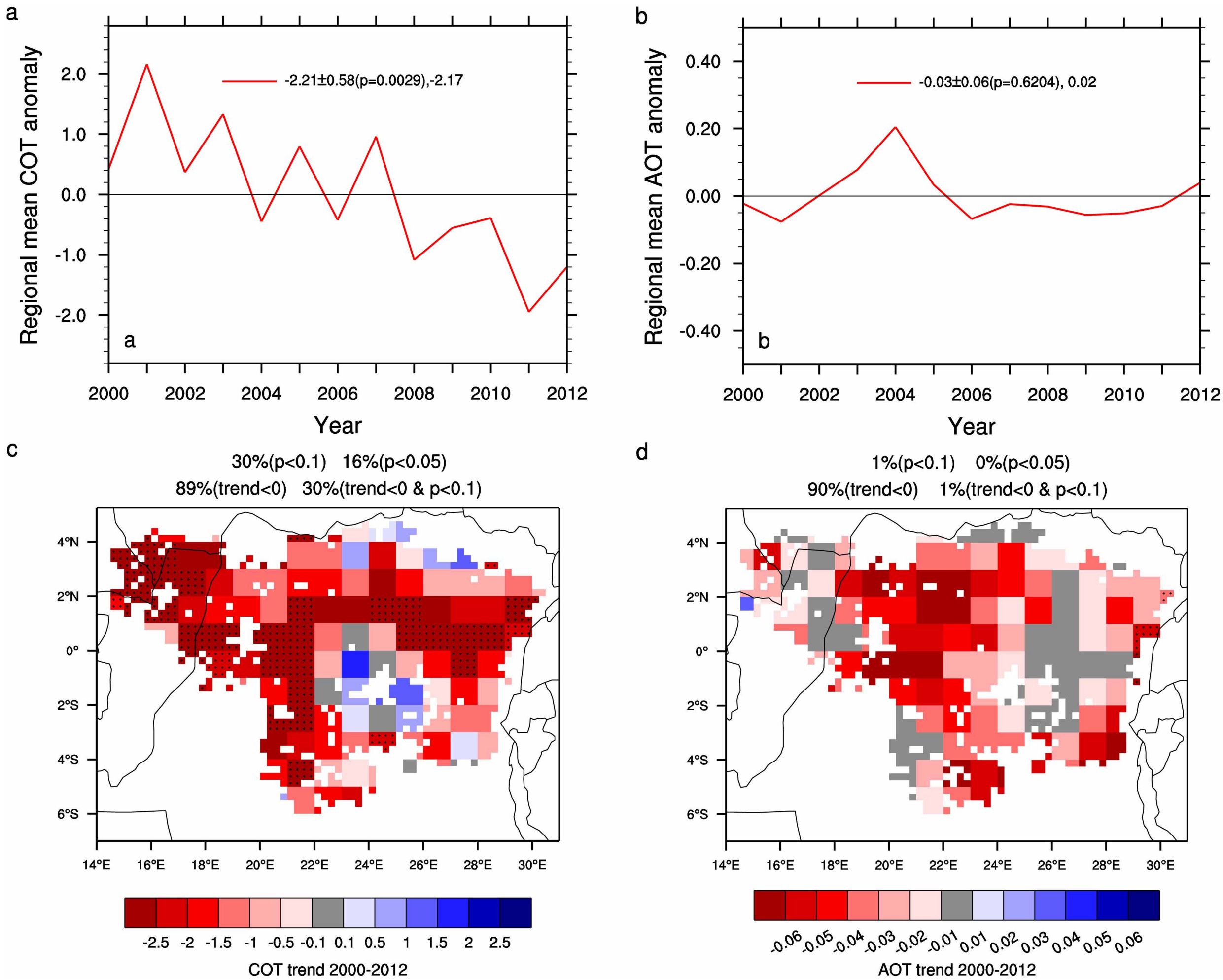
Collection 5 Percent Tree Cover. (Univ. Maryland, 2011); http://www.landcover.org/data/vcf.
34. King, M. D. *et al.* Cloud and aerosol properties, precipitable water and profiles of temperature and water vapor from MODIS. *IEEE Trans. Geosci. Rem. Sens.* **41,** 442–458 (2003).
35. Remer, L. A. *et al.* MODIS aerosol algorithm, products and validation. *J. Atmos. Sci.* **62,** 947–973 (2005).
36. Platnick, S. *et al.* The MODIS cloud products: algorithms and examples from Terra. *IEEE Trans. Geosci. Rem. Sens.* **41,** 459–473 (2003).
37. Hansen, M. C. *et al.* A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sens. Environ.* **112,** 2495–2513 (2008).
38. van Leeuwen, T. T. *et al.* Optimal use of land surface temperature data to detect changes in tropical forest cover. *J. Geophys. Res.* **116,** G02002 (2011).
39. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations. *Int. J. Climatol.* **34,** 623–642 (2013).
40. Landerer, F. W. & Swenson, S. C. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.* **48,** W04531 (2012).
41. Swenson, S. C. & Wahr, J. Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.* **33,** L08402 (2006).
42. Wielicki, B. A. *et al.* Clouds and the Earth's Radiant Energy System (CERES): an Earth observing system experiment. *Bull. Am. Meteorol. Soc.* **77,** 853–868 (1996).
43. Kirdyashev, K. P., Chukhlantsev, A. A. & Shutko, A. M. Microwave radiation of the Earth's surface in the presence of a vegetation cover. *Radio Eng. Electron. Phys.* **24,** 37–44 (1979).
44. Kerr, Y. H. & Njoku, E. G. A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. *IEEE Trans. Geosci. Remote Sens.* **28,** 384–393 (1990).
45. Jackson, T. J. & Schmugge, T. J. Vegetation effects on the microwave emission of soils. *Remote Sens. Environ.* **36,** 203–212 (1991).
46. Shi, J. C. *et al.* Microwave vegetation indices for short vegetation covers from satellite passive microwave sensor AMSR-E. *Remote Sens. Environ.* **112,** 4285–4300 (2008).
47. Jones, M. O., Jones, L. A., Kimball, J. S. & McDonald, K. C. Satellite passive microwave remote sensing for monitoring global land surface phenology. *Remote Sens. Environ.* **115,** 1102–1114 (2011).
48. Liu, Y. Y., de Jeu, R. A. M., McCabe, M. F., Evans, J. P. & van Dijk, A. I. J. M. Global long-term passive microwave satellite-based retrievals of vegetation optical depth. *Geophys. Res. Lett.* **38,** L18402 (2011).
49. Liu, Y. Y. *et al.* Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrol. Earth Syst. Sci.* **15,** 425–436 (2011).
50. Liu, Y. Y. *et al.* Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sens. Environ.* **123,** 280–297 (2012).
51. Halldor, B. & Venegas, S. A. *A Manual for EOF and SVD Analyses of Climate Data.* (Report No. 97-1, Centre for Climate and Global Change Research, McGill Univ., 1997).
52. Zhou, L., Tian, Y., Chen, H., Dai, Y. & Harris, R. A. Effects of topography on assessing wind farm impacts using MODIS data. *Earth Interact.* **17,** 1–18 (2013).



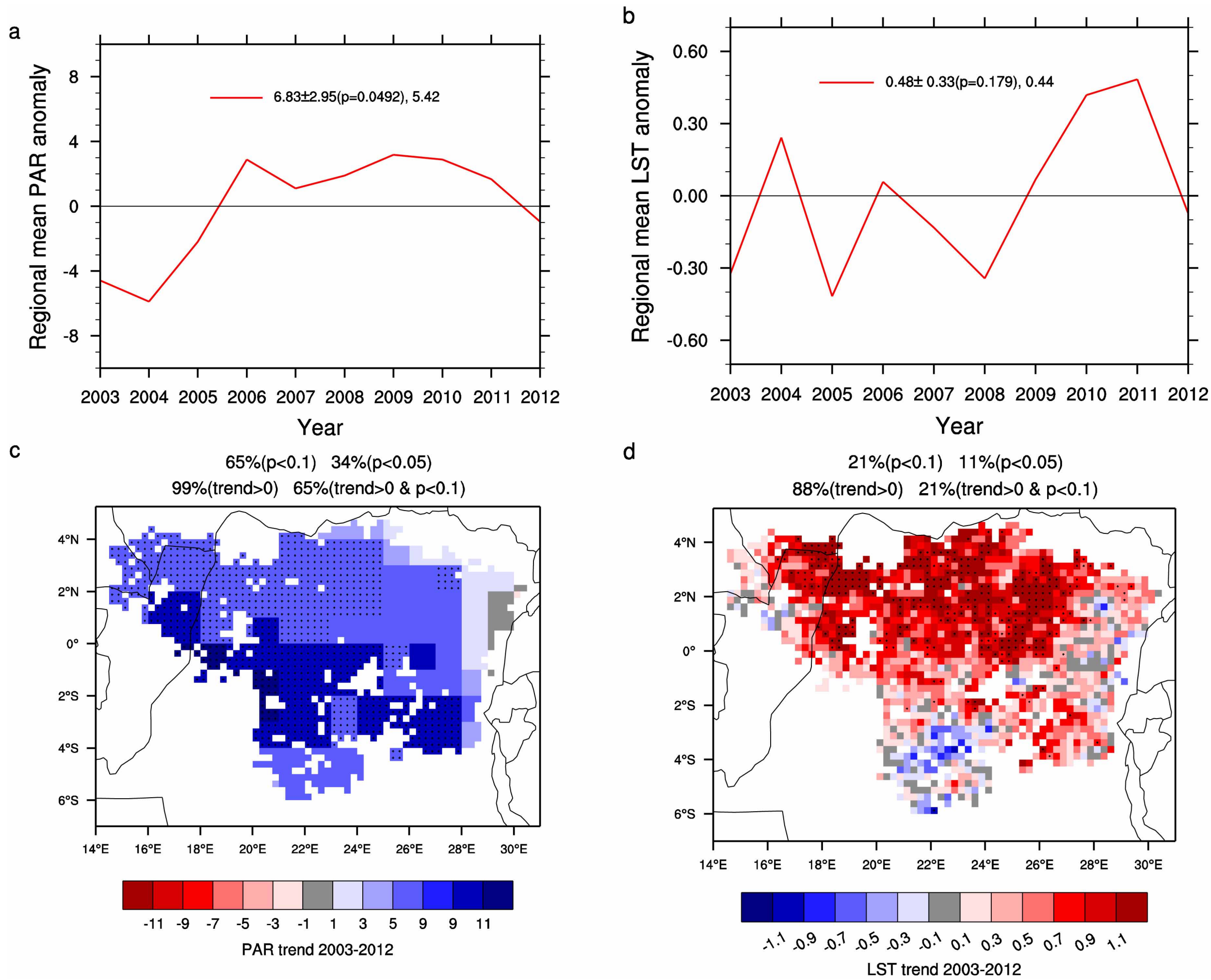
Extended Data Figure 1 | Statistics of MODIS data quality and MODIS high-quality EVI mask. **a–d**, Seasonal statistics of the total number of high-quality MODIS EVI composites over forested pixels at 0.05° resolution in central tropical Africa (7° S–7° N, 5–31 °E) for the period 2000–2012. For each season, there are up to 39 EVI composites (three composites per year multiplied by 13 years) for every pixel. **e**, The climatology of MODIS percentage forest cover at 0.05° resolution. **f**, The high-quality MODIS April–May–June EVI mask at 0.25° resolution over the intact Congo forest (6° S–5° N, 14–31° E) used in the analysis (see details in Methods).



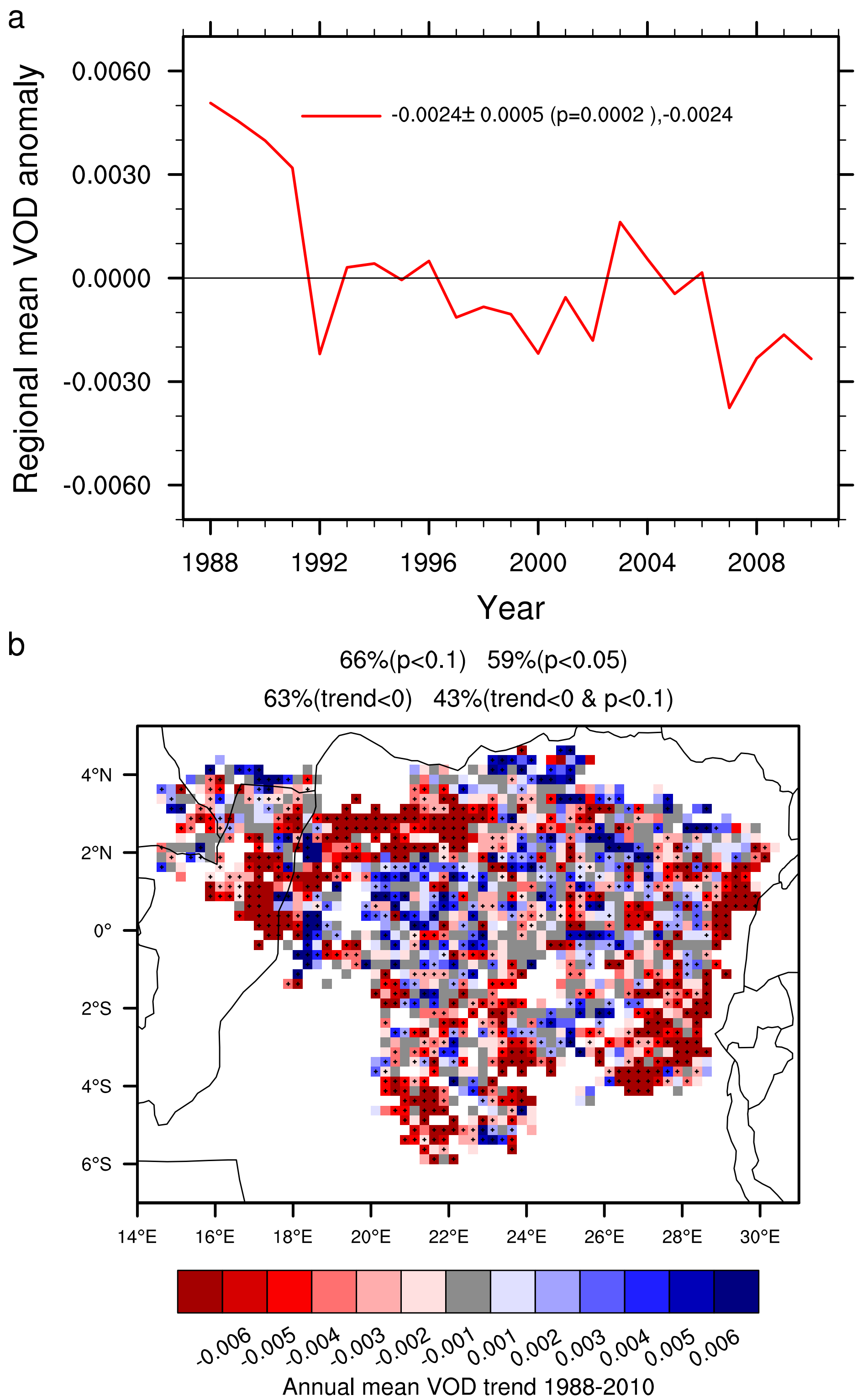
Extended Data Figure 2 | Spatial patterns of linear trends in April–May–June TWS (cm per decade) and correlation coefficients *R* between TWS and EVI for the period 2003–2012. a–c, As in Fig. 2e but for TWS from individual data processing centres (the University of Texas' Center for Space Research (CSR), NASA's Jet Propulsion Laboratory (JPL) and Germany's GeoForschungsZentrum (GFZ)). **d**, The topography of the Congo basin (http://en.wikipedia.org/wiki/File:Congo_Kinshasa_Topography.png). **e**, **f**, *R* between April–May–June EVI and ensemble-mean TWS in April–May–June (AMJ) and January–February–March (JFM). The significance level of *R* (its *P* value) is estimated using a two-tailed Student's *t*-test. Pixels with a plus symbol have a linear trend or an *R* that is statistically significant at $P < 0.1$. The percentages of pixels with trends or *R* at $P < 0.05$ and $P < 0.1$ and the percentages of pixels with negative trends or positive R over the study region are shown.



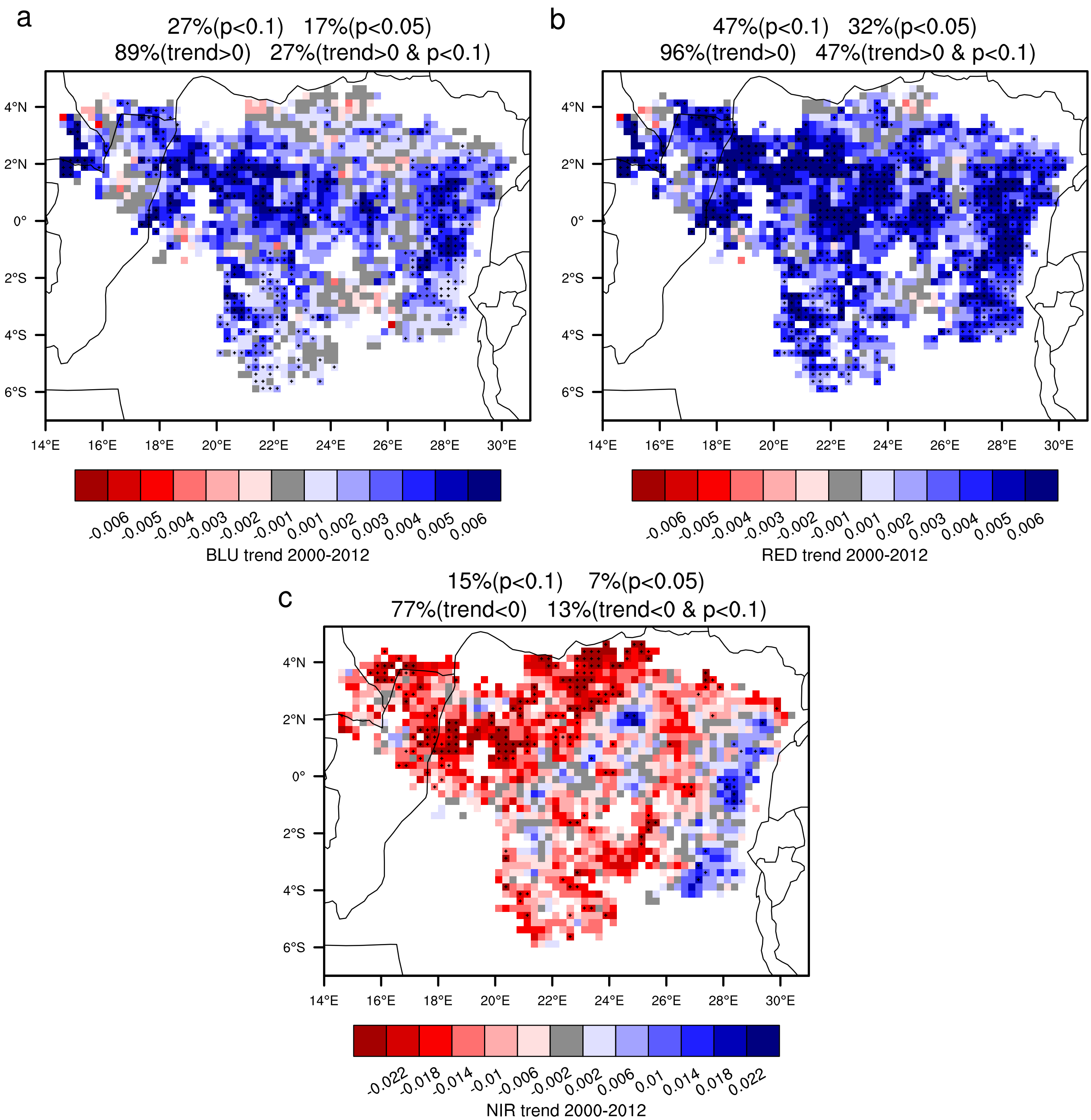
Extended Data Figure 3 | Regional mean anomalies and linear trends per decade for COT and AOT (unitless) in April–May–June for the period 2000–2012. **a**, **b**, As in Fig. 3a. **c**, **d**, As in Fig. 2a. The dramatic AOT increase in 2004 is due to volcanic eruptions of the mountains Nyamulagira and Nyiragongo, which are located on the eastern border of the study region, on 25 May 2004. However, if the year 2004 is excluded, the AOT changes little.



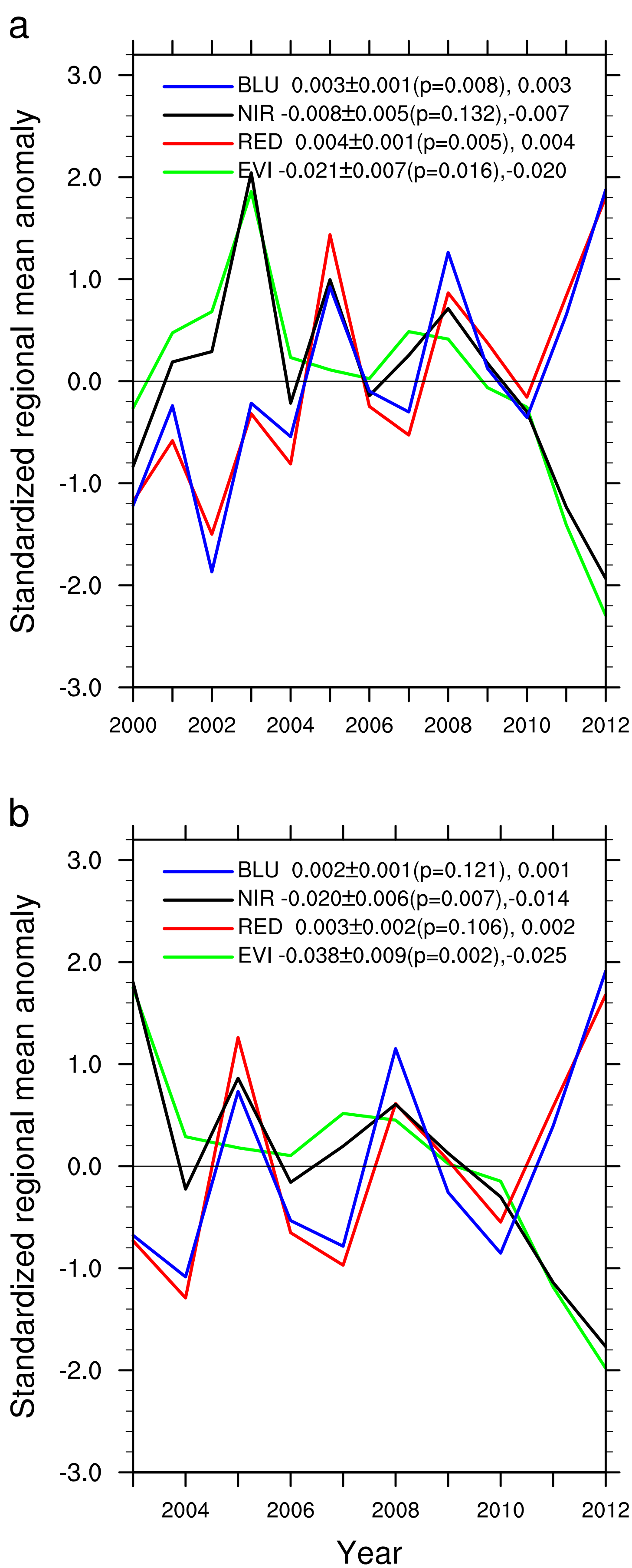
Extended Data Figure 4 | Regional mean anomalies and linear trends per decade for PAR and LST in April–May–June for the period 2003–2012. **a**, **c**, PAR (W m^{-2}); **b**, **d**, LST (°C); **a**, **b**, as in Fig. 3c; **c**, **d**, as in Fig. 2b.



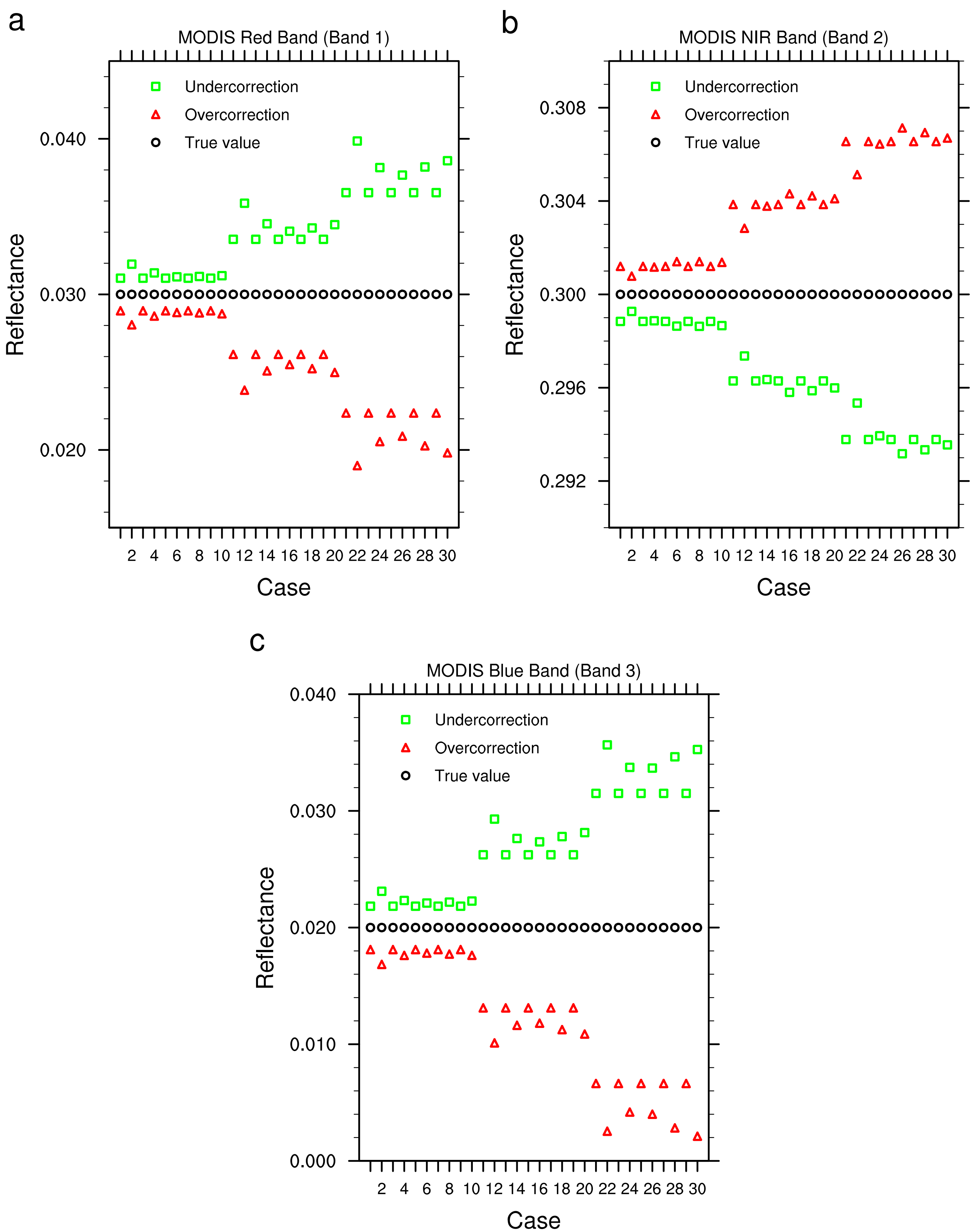
Extended Data Figure 5 | Annual mean VOD anomalies (unitless; a) and linear trends per decade (b). For the period 1988–2010 (as in Fig. 4).



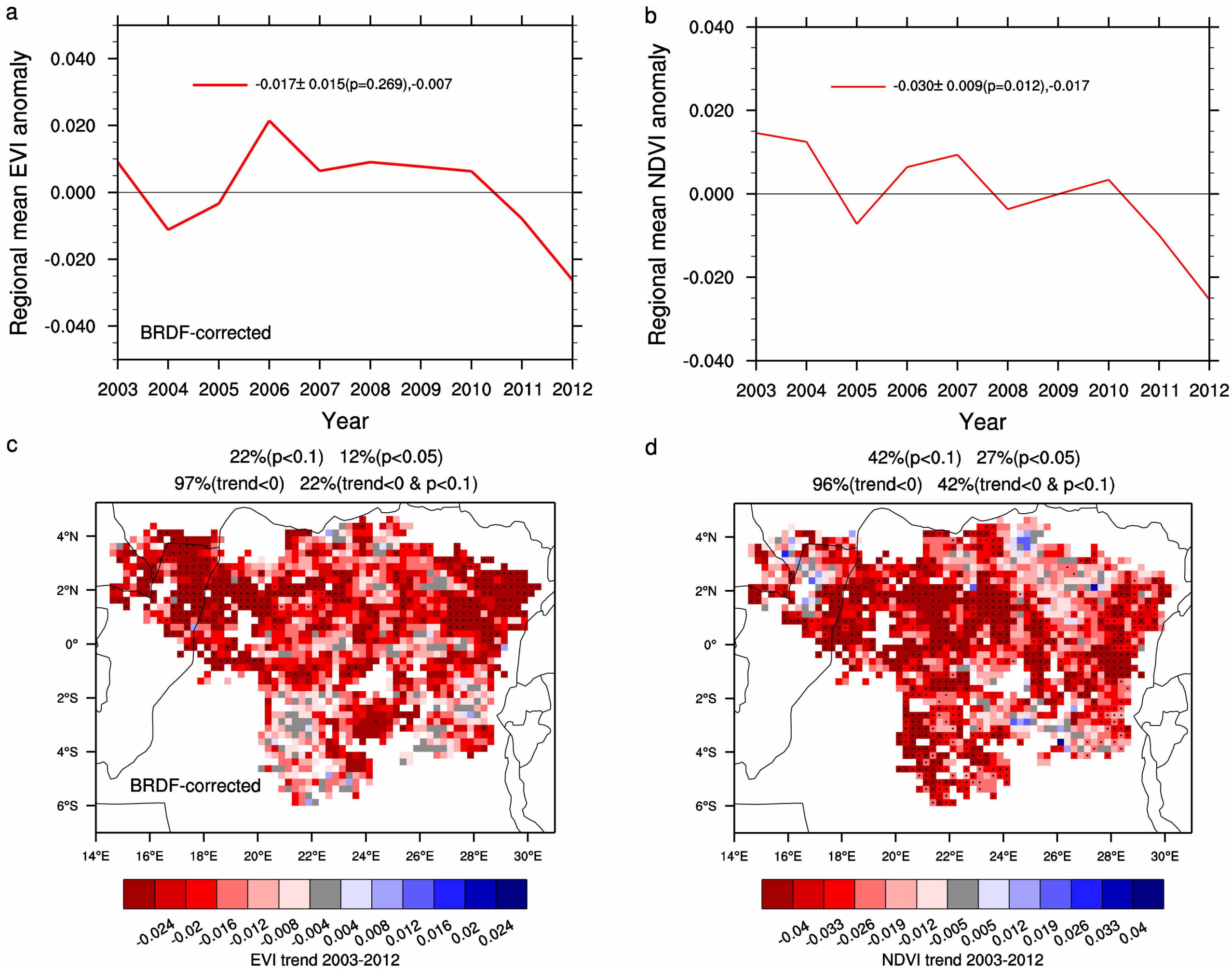
Extended Data Figure 6 | Spatial patterns of linear trends per decade in April–May–June for MODIS reflectance in the blue (BLU; a), red (RED; b) and near-infrared (NIR; c) spectral bands. For the period 2000–2012 (as in Fig. 2a).



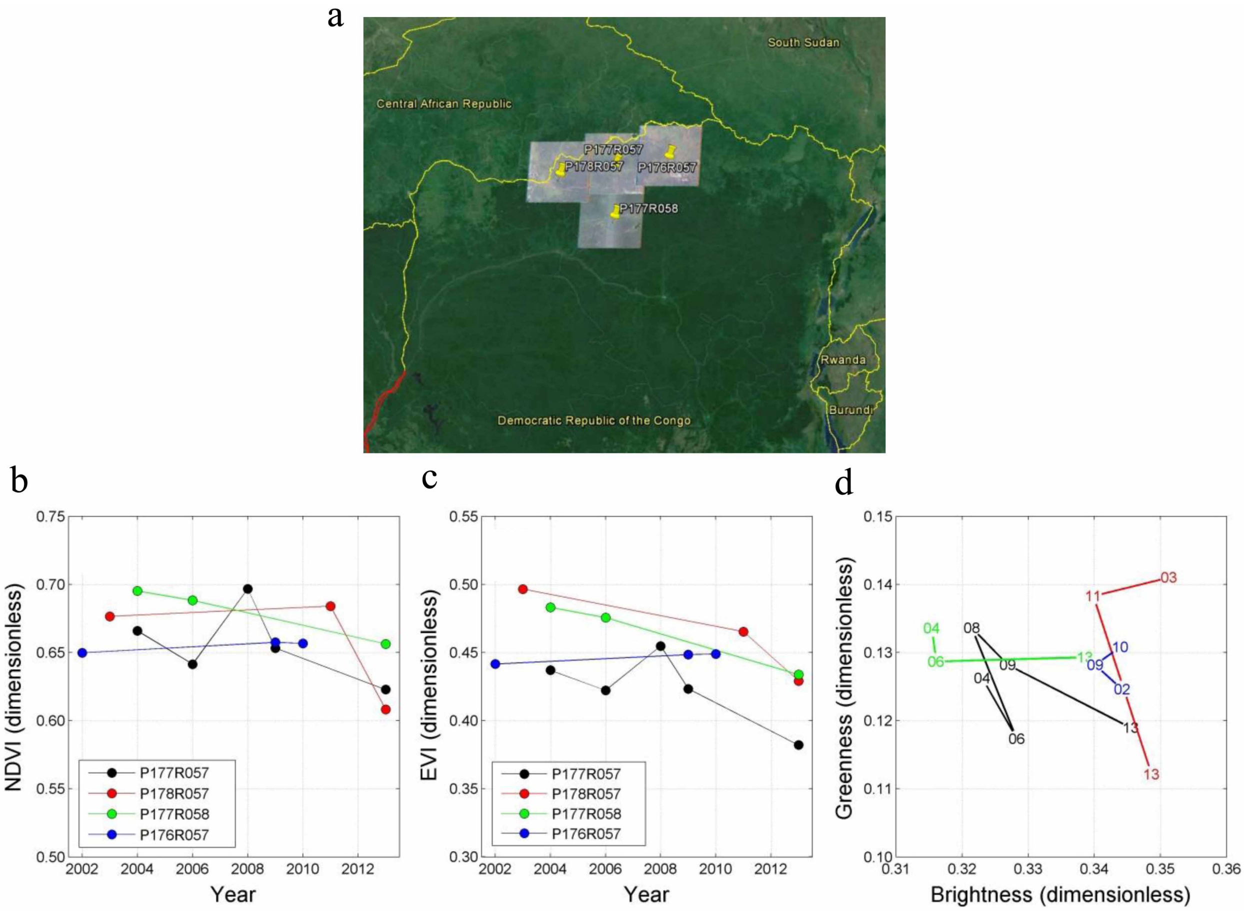
Extended Data Figure 7 | Regional mean anomalies for MODIS EVI and reflectance in the blue (BLU), red (RED) and near-infrared (NIR) spectral bands. For the period of 2000–2012 (**a**) (as in Fig. 3a) and 2003–2012 (**b**) (as in Fig. 3c).



Extended Data Figure 8 | Simulated surface reflectance values in the MODIS red (RED; a), near-infrared (NIR; b) and blue (BLU; c) bands using the 6S radiative transfer code for 25% overestimation or 25% underestimation of AOT. There are 30 cases (cases 1–10 correspond to a small AOT load, AOT = 0.1; cases 11–20 correspond to a medium AOT load, AOT = 0.3; cases 21–30 correspond to a large AOT load, AOT = 0.5) and the actual reflectance is 0.03, 0.3 and 0.02 in RED, NIR and BLU, respectively (see details in Supplementary Table 1).



Extended Data Figure 9 | Regional mean anomalies (unitless) and linear trends per decade for MODIS EVI and NDVI. For bidirectional reflectance distribution function (BRDF)-corrected EVI (**a**, **c**) calculated from MCD43C4 and for MODIS NDVI (**b**, **d**) from MOD13C2 (as in Extended Data Fig. 4).



Extended Data Figure 10 | Temporal dynamics of vegetation for four Landsat 7 ETM+ scenes. **a**, Locations of the Landsat scenes (P176R057, P177R057, P178R057 and P177R058). **b**, Mean temporal variations of NDVI for cloud-free pixels with NDVI ≥ 0.5 in the first of the image time series. **c**, Mean temporal variations of EVI for the same pixels as in **b**. **d**, Mean temporal trajectory of vegetation in the brightness–greenness space of the Tasseled Cap transformation (see details in Supplementary Information section D). A decrease in greenness associated with an increase in brightness signifies forest degradation.

LETTER

doi:10.1038/nature13176

Identification of genomic alterations in oesophageal squamous cell cancer

Yongmei Song[1]*, Lin Li[2]*, Yunwei Ou[1,3]*, Zhibo Gao[2]*, Enmin Li[4]*, Xiangchun Li[2]*, Weimin Zhang[1], Jiaqian Wang[2], Liyan Xu[5], Yong Zhou[2], Xiaojuan Ma[1], Lingyan Liu[1], Zitong Zhao[1], Xuanlin Huang[2], Jing Fan[1], Lijia Dong[1], Gang Chen[2], Liying Ma[1], Jie Yang[2], Longyun Chen[2], Minghui He[2], Miao Li[2], Xuehan Zhuang[2], Kai Huang[2], Kunlong Qiu[2], Guangliang Yin[2], Guangwu Guo[2], Qiang Feng[2], Peishan Chen[2], Zhiyong Wu[6], Jianyi Wu[4], Ling Ma[1], Jinyang Zhao[2], Longhai Luo[2], Ming Fu[1], Bainan Xu[3], Bo Chen[5], Yingrui Li[2], Tong Tong[1], Mingrong Wang[1], Zhihua Liu[1], Dongxin Lin[1], Xiuqing Zhang[2], Huanming Yang[2], Jun Wang[2] & Qimin Zhan[1]

Oesophageal cancer is one of the most aggressive cancers and is the sixth leading cause of cancer death worldwide[1]. Approximately 70% of global oesophageal cancer cases occur in China, with oesophageal squamous cell carcinoma (ESCC) being the histopathological form in the vast majority of cases (>90%)[2,3]. Currently, there are limited clinical approaches for the early diagnosis and treatment of ESCC, resulting in a 10% five-year survival rate for patients. However, the full repertoire of genomic events leading to the pathogenesis of ESCC remains unclear. Here we describe a comprehensive genomic analysis of 158 ESCC cases, as part of the International Cancer Genome Consortium research project. We conducted whole-genome sequencing in 17 ESCC cases and whole-exome sequencing in 71 cases, of which 53 cases, plus an additional 70 ESCC cases not used in the whole-genome and whole-exome sequencing, were subjected to array comparative genomic hybridization analysis. We identified eight significantly mutated genes, of which six are well known tumour-associated genes (*TP53*, *RB1*, *CDKN2A*, *PIK3CA*, *NOTCH1*, *NFE2L2*), and two have not previously been described in ESCC (*ADAM29* and *FAM135B*). Notably, *FAM135B* is identified as a novel cancer-implicated gene as assayed for its ability to promote malignancy of ESCC cells. Additionally, *MIR548K*, a microRNA encoded in the amplified 11q13.3-13.4 region, is characterized as a novel oncogene, and functional assays demonstrate that *MIR548K* enhances malignant phenotypes of ESCC cells. Moreover, we have found that several important histone regulator genes (*MLL2* (also called *KMT2D*), *ASH1L*, *MLL3* (*KMT2C*), *SETD1B*, *CREBBP* and *EP300*) are frequently altered in ESCC. Pathway assessment reveals that somatic aberrations are mainly involved in the Wnt, cell cycle and Notch pathways. Genomic analyses suggest that ESCC and head and neck squamous cell carcinoma share some common pathogenic mechanisms, and ESCC development is associated with alcohol drinking. This study has explored novel biological markers and tumorigenic pathways that would greatly improve therapeutic strategies for ESCC.

To discover genomic alterations in human ESCC, we collected tumour and matched normal DNA samples from 158 ESCC patients in the Chaoshan District of Guangdong Province, an area of high ESCC prevalence in China, and performed whole-genome sequencing (WGS >30×), whole-exome sequencing (WES >100×), and array comparative genomic hybridization (a-CGH) analyses (Supplementary Tables 1–3 and Extended Data Fig. 1). We identified 7,182 single nucleotide variants (SNVs) and 97 insertions and deletions (indels) (Supplementary Tables 4 and 5), with a median of 2.60 per megabase (Mb) (range 0.03–7.79) in protein-coding regions of 17 WGS and 71 WES cases (Supplementary Table 6). A total of 318 out of 325 SNVs and 16 out of 21 indels were verified as somatic mutations (Supplementary Table 7). We also identified 25,159 copy number alterations (CNAs) (median = 101.5; range 2–2,898) (Supplementary Table 8) and obtained 58 significant CNAs in 17 WGS and 123 a-CGH samples (Supplementary Table 9). In addition, we identified 1,737 structural variants (SVs) (median = 92, range 7–238) in 17 WGS samples (Supplementary Table 10), and verified 16 out of 18 breakpoints (Supplementary Table 11).

The somatic mutational rate for ESCC is higher than breast carcinoma and glioblastoma multiforme, but less than head and neck squamous cell carcinoma (HNSCC), oesophageal adenocarcinoma (EAC) and lung squamous cell carcinoma (SQCC)[4–6]. C•G>T•A transitions were the most common mutations, followed by C•G>G•C transversions, which is consistent with previous reports on ESCC[7]. Compared with EAC[5], HNSCC[6] and lung SQCC[4], the overall pattern of the mutation spectrum in ESCC was similar to that of HNSCC, but different from that of lung SQCC, which is dominated by C•G>A•T transversions. Although the most common mutations in EAC were also C•G>T•A transitions, the second most frequent mutations were T•A>G•C transversions (Extended Data Fig. 2). Through hierarchical clustering analysis, we found that ESCC and HNSCC mutation spectra were intermingled,

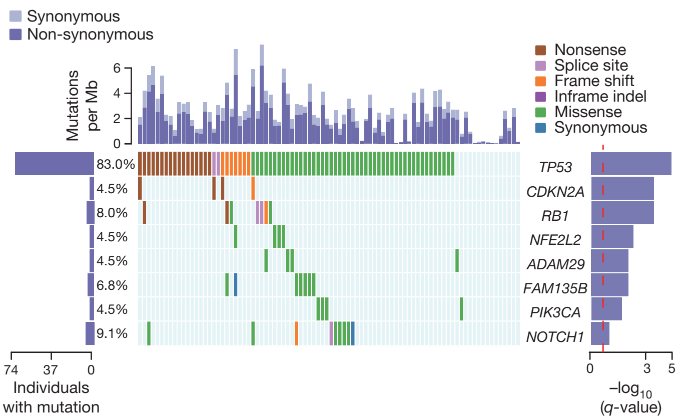


Figure 1 | Significantly mutated genes in ESCC. Top: rate of synonymous and non-synonymous mutations in 88 tumours, expressed in number of mutations per megabase (Mb) of covered target sequence. The middle panel shows the matrix of mutations in significantly mutated genes coloured by the types of mutation. Each column denotes an individual tumour, and each row represents a gene. The left panel shows the number of mutations in each gene. Percentages represent the fraction of tumours harbouring mutations in the corresponding gene. The right panel shows significantly mutated genes ranked by q-value.

[1]State Key Laboratory of Molecular Oncology, Cancer Institute and Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China. [2]BGI-Shenzhen, Shenzhen 518083, Guangdong 518083, China. [3]Department of Neurosurgery, Chinese PLA General Hospital, Beijing 100853, China. [4]Department of Biochemistry and Molecular Biology, The Key Laboratory of Molecular Biology for High Cancer Incidence Coastal Chaoshan Area, Shantou University Medical College, Shantou 515041, Guangdong, China. [5]Institute of Oncologic Pathology, Shantou University Medical College, Shantou 515041, Guangdong, China. [6]Department of Tumor Surgery, Shantou Central Hospital, Affiliated Shantou Hospital of Sun Yat-sen University, Shantou 515041, Guangdong, China.
*These authors contributed equally to this work.

whereas lung SQCCs and EACs were clearly distinctive (Extended Data Fig. 3).

Context-specific mutation spectrum analysis of 88 ESCCs yielded three distinct clusters[8]. Cluster 1 was significantly enriched in non-drinking patients (never consumed alcohol) ($P = 0.003$, Fisher's exact test), and marked for better prognosis compared with patients of cluster 2 by Kaplan–Meier analysis ($P = 0.022$, log-rank test) (Supplementary Tables 12 and 13 and Extended Data Fig. 4). These results indicate that the risk of ESCC is correlated with alcohol consumption.

We identified eight significantly mutated genes (Methods in Supplementary Information), including six well-known ESCC-implicated genes (*TP53*, *RB1*, *CDKN2A*, *PIK3CA*, *NOTCH1* and *NFE2L2*), one tumour-associated gene (*ADAM29*) that has not previously been described in ESCC, and one gene (*FAM135B*) that has not been linked to cancer previously (Fig. 1). *TP53* somatic mutations were identified in 83.0% of samples, mainly located in the DNA-binding domain (85%, 68 of 80) (Extended Data Fig. 5). *ADAM29*, a member of the ADAM family that is frequently mutated in melanoma[9], contained four missense mutations, with two of them in the same residue (p.Q814H), probably forming a mini-hotspot.

FAM135B (family with sequence similarity 135, member B) was mutated in 6.8% (6 of 88) of cases and was associated with poor prognosis in ESCC ($P = 0.026$, log-rank test) (Extended Data Fig. 6 and Supplementary Table 14). Additionally, we detected *FAM135B* amplifications in 25.0% (35 of 140) of cases. We assessed *FAM135B* expression in nine ESCC cell lines (containing wide-type *FAM135B*) and an

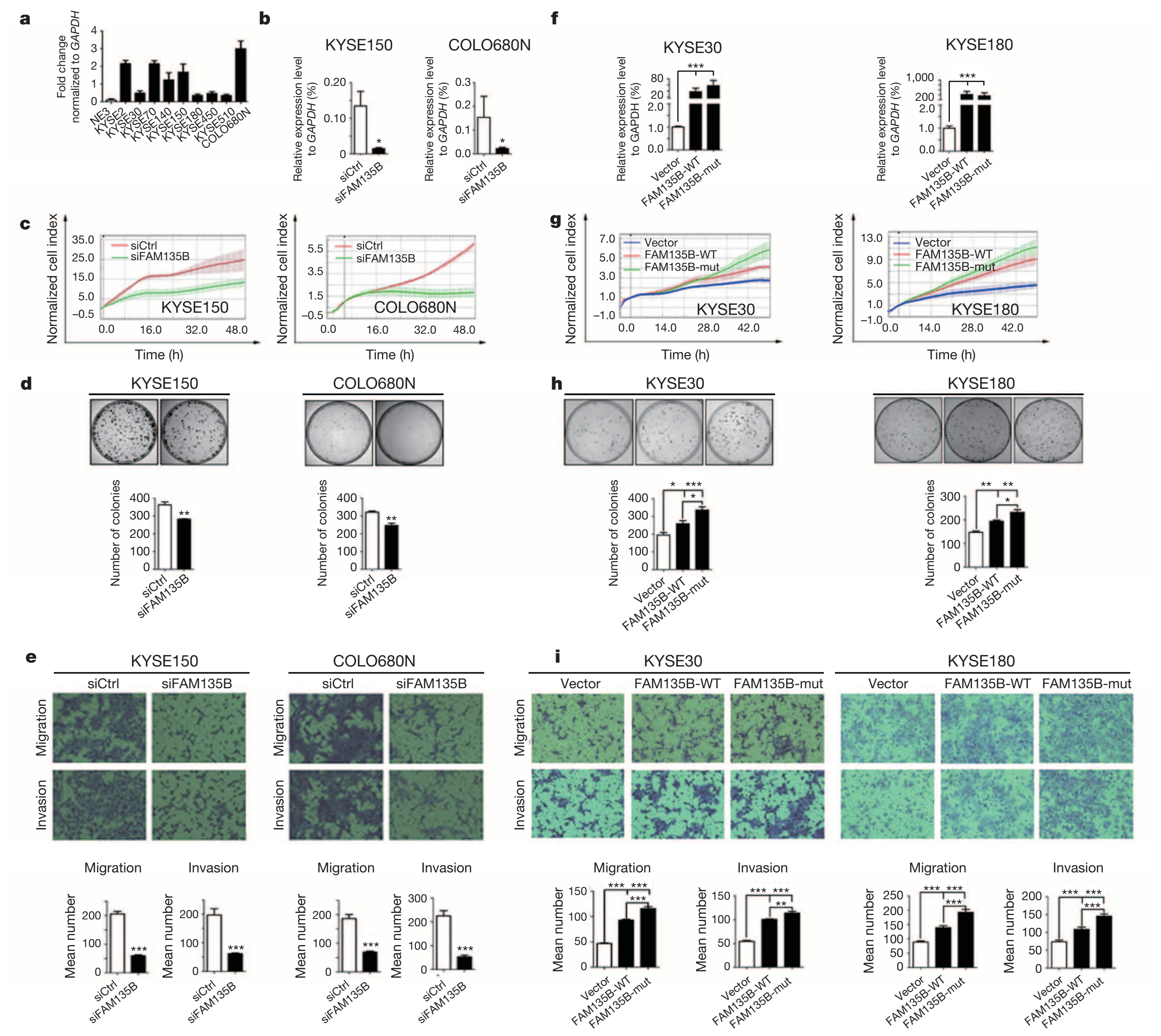


Figure 2 | *FAM135B* positively modulates ESCC cellular malignant phenotypes. **a**, The *FAM135B* levels in 9 ESCC cell lines and an immortalized normal oesophageal squamous cell line (NE3) were analysed by real-time PCR. **b**, Depleted *FAM135B* levels after siRNA knockdown in KYSE150 and COLO680N cell lines were measured by real-time PCR. **c**–**e**, Knockdown of *FAM135B* reduced cell proliferation, colony formation, migration and invasion. **f**, Ectopic expression of wild-type and mutant *FAM135B* in KYSE30 and KYSE180 cell lines was analysed by real-time PCR. **g**–**i**, Overexpression of wild-type *FAM135B* substantially enhanced cell proliferation, colony formation, migration and invasion, whereas *FAM135B* mutant had a greater effect in increasing malignant phenotypes of ESCC cells. Representative images (top) and quantification (bottom) are shown in colony formation and transwell (migration or invasion) assays. All experiments were performed at least three times and data statistically analysed by two-sided *t*-test. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ versus control. FAM135B-wt, wild-type *FAM135B*; FAM135B-mut, mutant *FAM135B*. Error bars indicate s.e.m.

immortalized normal oesophageal cell line, and found that *FAM135B* was highly expressed in all ESCC cell lines compared with normal control (Fig. 2a). We then knocked down *FAM135B* through short interfering RNA (siRNA) in KYSE150 and COLO680N cells, which express relatively higher levels of *FAM135B*, and observed that depletion of *FAM135B* attenuated cellular malignant phenotypes including cell growth, colony formation, migration and invasion (Fig. 2b–e). We next constructed wide-type and mutant (p.S165P) *FAM135B* vectors and transiently transfected them into KYSE30 and KYSE180 cells, which express lower levels of *FAM135B*. Consistently, wild-type *FAM135B* significantly enhanced malignant phenotypes, whereas the *FAM135B* mutant showed stronger capability of promoting malignant phenotypes compared to wild-type *FAM135B* (Fig. 2f–i). These findings indicate that *FAM135B* contributes to the development of ESCC and may serve as a prognostic marker or therapeutic target for ESCC.

Notably, we detected frequent non-silent mutations in 48 histone-modification-related genes in 53.4% of cases (Supplementary Table 15). The most frequent alterations were observed in histone-H3-modifying

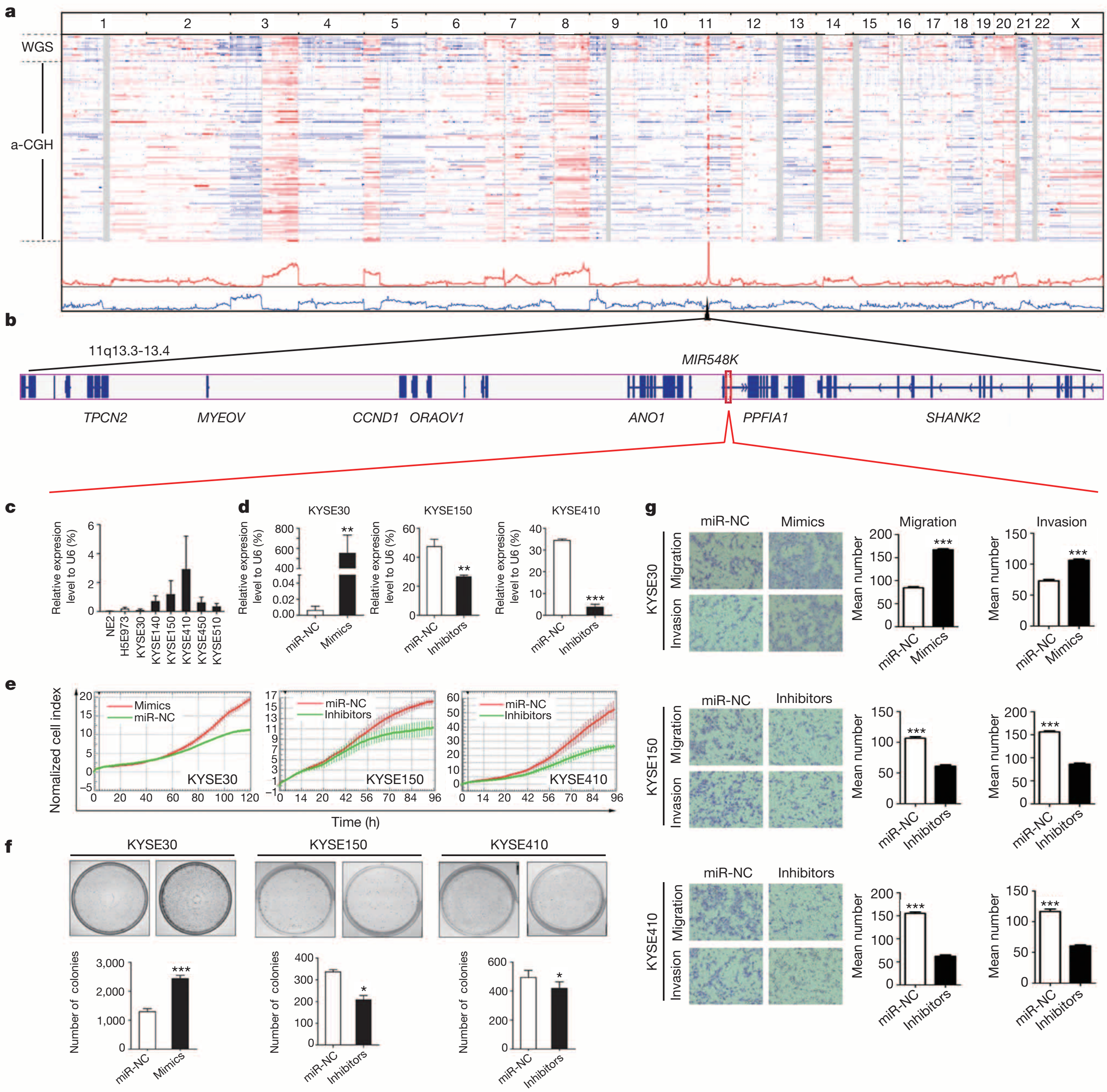


Figure 3 | Landscape of genomic copy number alterations in ESCC and oncogenic *MIR548K* identified from significantly amplified region. **a**, Landscape of genomic copy number alterations in 123 a-CGH cases and 17 WGS cases. Copy number gains are shown in shades of red, and shades of blue for copy number losses. **b**, The location of *MIR548K* in the significantly amplified 11q13.3-13.4 region. **c–g**, *MIR548K* positively modulates ESCC cell malignant phenotypes. **c**, The levels of *MIR548K* in six ESCC cell lines and two immortalized normal oesophageal cell lines were analysed by real-time PCR. **d**, Ectopic expression and knockdown of *MIR548K* in ESCC cell lines (KYSE30, KYSE150 and KYSE410) were measured by real-time PCR. **e**, Growth curves show that *MIR548K* greatly promotes cell proliferation in KYSE30, KYSE150 and KYSE410 cells. **f**, *MIR548K* enhanced colony formation of ESCC cells. Representative pictures (top) and quantitative analysis (bottom) of colony formation assays. **g**, *MIR548K* increased cellular motility of KYSE30, KYSE150 and KYSE410 cells. Representative pictures (left) and quantitative data (right) of transwell (migration or invasion) assays. All experiments were performed at least three times and data were statistically analysed by two-sided *t*-test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ versus control. Error bars indicate s.e.m.

lysine methyltransferases (21.6% of cases) (Supplementary Table 16), in which four genes were mutated in more than two samples: *MLL2*, *ASH1L*, *MLL3* and *SETD1B* (Extended Data Fig. 7). Another two genes encoding histone H3 lysine acetyltransferases, *CREBBP* and *EP300*, also displayed a high frequency of somatic mutation.

We identified large-scale chromosome amplifications at 3q, 5p, 8q, 12p, 20p and 20q and deletions at 3p, 4q, 9p, 13q, 18q, 19p and 21q (Fig. 3a), which further support the findings by ref. 10. Compared with EAC, HNSCC and lung SQCC, the CNA landscape in ESCC was similar to that in HNSCC (gain of 1q, 7q and 11p; loss of 9p and 10q), but different from EAC (gain of 8q, 7p and 12p; loss of Xp, 21q and 22q) and lung SQCC (gain of 2p, 5p, 17q, 19q, 22q and Xp; loss of 1p, 4q and 13q) (Extended Data Fig. 8). Furthermore, we analysed shared properties between ESCC and HNSCC, and found that loss of *RAF1* and *RASA1*, and amplification of *EGFR*, *MYC*, *KRAS* and *AKT1*, exhibited similar frequencies, suggesting that MAPK pathway activation might be a common event for ESCC and HNSCC. Additionally, similar CNA profiles of many genes in cell cycle and JAK–STAT pathways were detected in both ESCC and HNSCC (Extended Data Fig. 9). These findings, together with the mutation spectrum analysis, suggest that ESCC and HNSCC pathogenesis may share some common characteristics and analogous therapeutic approaches could be considered for them.

Meanwhile, 1,325 genes in 43 significant amplified regions and 229 genes in 15 significant deleted regions were characterized. Among these 58 focal CNAs, 6 significantly amplified regions in ESCC were associated with regional lymph node involvement, including 11q13.3-13.4, 8q24.3 and 14q32.33 ($P < 0.001$, $P = 0.009$ and $P = 0.002$, respectively, Fisher's exact test) (Supplementary Table 17). The 11q13.3-13.4 region has been reported to be amplified in many human cancers[11,12], and in this region we identified one microRNA, *MIR548K*, that has never been associated with human cancers (Fig. 3b). Elevated expression of *MIR548K* was detected in ESCC cell lines compared with normal oesophageal cells. Overexpression of *MIR548K* substantially enhanced the malignant phenotypes of ESCC, whereas depletion of *MIR548K* via an inhibitor suppressed ESCC cellular growth and mobility (Fig. 3c–g).

We detected 890 genes that harboured SVs in 17 WGS-analysed samples, 25 of which were characterized in three or more cases (Supplementary Table 18 and Extended Data Fig. 10). The most frequently altered gene was *KCNB2*, which encodes a protein mediating the voltage-dependent potassium ion permeability of excitable membranes and is associated with glioma[13]. We also performed analysis on several viruses including human papilloma virus, hepatitis B virus and human herpesvirus[14–16], and found no viral integration in the ESCC genome (Supplementary Table 19).

We integrated all SNVs and CNAs from the 158 ESCC cases and determined genomic alterations of several important pathways, including Wnt, cell cycle, Notch, RTK–Ras and AKT pathways (Fig. 4) (Supplementary Table 20).

Altered genes in the Wnt pathway were detected in 86.4% of cases, including mutations in *CTNNB1* (1.1%) and *SFRP4* (3.4%). The *SFRP4* mutations are located in the CRD domain, which is critical for SFRP4 to antagonize the Wnt pathway[17]. Many genes that antagonize *AXIN2* (a canonical Wnt suppressor), including *DAAM2*, *DVL3*, *LRP5* and *LRP6*, displayed mutations and amplifications in 46.2% of cases. The *YAP1* transcriptional complexes, which are essential for β-catenin-driven tumorigenesis[18], were altered in 6.4% of samples, and antiapoptotic genes *BIRC5* and *BCL2L1* were amplified in 7.1% and 15.0% of cases, respectively.

Genetic alterations of cell cycle pathway in ESCC mainly exhibited distinct defects in G1/S transition control, as demonstrated by amplifications of *CCND1*, *CDK4*, *CDK6*, *E2F1* and *MDM2* in 62.9% of cases, and mutations or deletions of *RB1*, *CDKN2A*, *CHEK1*, *CHEK2* and *TP53* in 66.5% of cases. In addition to *EGFR* amplification, *EGFR* downstream signalling cascades such as the RTK–Ras and AKT pathways also displayed genetic alterations in 78.6% of cases, including *KRAS*, *MRAS*, *RAF1*, *AKT1*, *SOS1*, *SOS2* and *PIK3CA* mutations and amplifications in 50.6% of cases.

The Notch pathway was recurrently mutated in 35.2% of cases. *NOTCH1* showed mutations in eight cases, and four mutations in *NOTCH2* and two mutations in *NOTCH3* were also detected. The mutation and amplification of *NOTCH1*, *NOTCH2* and *NOTCH3* were present in 16.4% of cases.

In potential therapeutic target analysis, mutations of validated and candidate therapeutic targeted genes[19,20] were discovered in 42% of samples and the most frequently mutated target was PI(3)K (class I). *EGFR* displayed one nonsense mutation in p.E665X, whereas its amplification was detected in 24.3% of samples. We also analysed amplification data and identified 136 genes harbouring potentially drugable alterations in 117 cases, including some novel therapeutic targets such as *PSMD2*, *RARRES1*, *SRC*, *GSK3B* and *SGK3* (Supplementary Table 21).

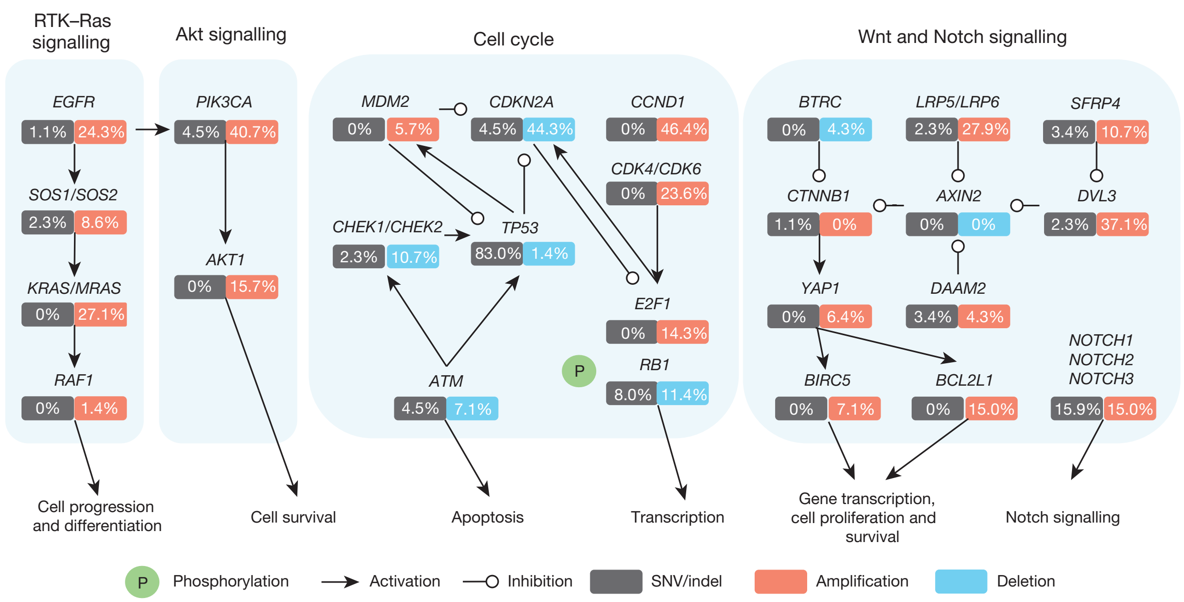


Figure 4 | Somatically altered pathways in ESCC. From right to left are Notch, Wnt, cell cycle, Akt and RTK–Ras signalling pathways. Alteration frequencies are expressed as the percentage of cases (somatic mutation cases = 88; CNA cases = 140). SNV/indel is highlighted in brown, amplification in red, and deletion in blue.

Additionally, PI(3)K/MTOR and ERK pathways were identified as important potential therapeutic targets in ESCC (Supplementary Table 22).

This study represents a comprehensive characterization of genomic alterations in ESCC, and provides insights into the genetic mechanism(s) of ESCC tumorigenesis. These findings enable us to determine further the biological and therapeutic significance of the newly discovered mutated and amplified genes, which may ultimately lead to the development of effective diagnostic and therapeutic approaches for ESCC.

METHODS SUMMARY

As part of the International Cancer Genome Consortium (ICGC) oesophageal cancer project (http://icgc.org/icgc/cgp/72/371/1001734), a total of 158 tumour and matched normal (blood) samples were obtained from patients with documented informed consent through the institutional review board. In the WGS and WES, all samples had library construction and sequenced on the Illumina Hiseq 2000 platform. Paired-end reads were aligned to NCBI build37 using BWA and duplicated reads were marked by Picard[21]. We detected SNVs and indels by Varscan, GATK and in-house filter pipeline, and used Annovar[22–24] for annotation. For WGS, CNAs were predicted by SegSeq and smoothed using DNAcopy[25,26]. SegMNT and DNAcopy were used to process a-CGH data[26,27]. Candidate somatic SVs were identified by CREST[28]. A subset of sequencing variants was validated by PCR, mass-spectrometric genotyping or Sanger sequencing. Sequencing and array-CGH data have been deposited to EGA (EGAS00001000709) and GEO (GSE54995). Detailed methods are provided in Supplementary Information.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 May 2013; accepted 25 February 2014.

Published online 16 March 2014.

1. Kamangar, F., Dores, G. M. & Anderson, W. F. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J. Clin. Oncol.* **24,** 2137–2150 (2006).
2. Xu, Y., Yu, X., Chen, Q. & Mao, W. Neoadjuvant versus adjuvant treatment: which one is better for resectable esophageal squamous cell carcinoma? *World J. Surg. Oncol.* **10,** 173 (2012).
3. Zhang, S. W. *et al.* An analysis of incidence and mortality of esophageal cancer in China, 2003–2007. *China Cancer* **21,** 241–247 (2012).
4. The Caner Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).
5. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature Genet.* **45,** 478–486 (2013).
6. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333,** 1157–1160 (2011).
7. Agrawal, N. *et al.* Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov.* **2,** 899–905 (2012).
8. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150,** 1107–1120 (2012).
9. Wei, X. *et al.* Analysis of the disintegrin-metalloproteinases family reveals ADAM29 and ADAM7 are often mutated in melanoma. *Hum. Mutat.* **32,** E2148–E2175 (2011).
10. Bandla, S. *et al.* Comparative genomics of esophageal adenocarcinoma and squamous cell carcinoma. *Ann. Thorac. Surg.* **93,** 1101–1106 (2012).
11. Ying, J. *et al.* Genome-wide screening for genetic alterations in esophageal cancer by aCGH identifies 11q13 amplification oncogenes associated with nodal metastasis. *PLoS ONE* **7,** e39797 (2012).
12. Komatsu, Y. *et al.* TAOS1, a novel marker for advanced esophageal squamous cell carcinoma. *Anticancer Res.* **26,** 2029–2032 (2006).
13. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321,** 1807–1812 (2008).
14. McLaughlin-Drubin, M. E., Meyers, J. & Munger, K. Cancer associated human papillomaviruses. *Curr. Opin. Virol.* **2,** 459–466 (2012).
15. Arzumanyan, A., Reis, H. M. & Feitelson, M. A. Pathogenic mechanisms in HBV- and HCV-associated hepatocellular carcinoma. *Nature Rev. Cancer* **13,** 123–135 (2013).
16. Panagiotakis, G. I. *et al.* Association of human herpes, papilloma and polyoma virus families with bladder cancer. *Tumour Biol.* **34,** 71–79 (2013).
17. Longman, D., Arfuso, F., Viola, H. M., Hool, L. C. & Dharmarajan, A. M. The role of the cysteine-rich domain and netrin-like domain of secreted frizzled-related protein 4 in angiogenesis inhibition *in vitro*. *Oncol. Res.* **20,** 1–6 (2012).
18. Rosenbluh, J. *et al.* β-Catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. *Cell* **151,** 1457–1473 (2012).
19. Forde, P. M. & Kelly, R. J. Genomic alterations in advanced esophageal cancer may lead to subtype-specific therapies. *Oncologist* **18,** 823–832 (2013).
20. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483,** 570–575 (2012).
21. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
22. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22,** 568–576 (2012).
23. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
24. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
25. Chiang, D. Y. *et al.* High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nature Methods* **6,** 99–103 (2009).
26. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23,** 657–663 (2007).
27. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5,** 557–572 (2004).
28. Wang, J. *et al.* CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nature Methods* **8,** 652–654 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work is supported by the funding from the National High Technology Research and Development Program of China (863 program no. 2012AA02A209 and 2012AA02A503), National Natural Science Foundation Fund (81021061), Guangdong Innovative Research Team Program (2009010016), the National Natural Science Foundation of China-GuangDong Joint Fund (U0932001), and the National Key Basic Research Program of China (973 program no. 2011CB911004, 2009CB521801 and 2012CB526608). The ESCC cell lines (KYSE30, KYSE70, KYSE180, KYSE410, KYSE450, KYSE140, COLO680N and KYSE510) were provided by Y. Shimada of Kyoto University. We also acknowledge International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) for sharing the EAC, HNSCC and lung SQCC data.

Author Contributions Q.Z. and Y.S. contributed to the design of the project and Q.Z. also mainly contributed to writing the manuscript. E.L., L.X., Z.W., Jianyi Wu and B.C. provided clinical samples and relevant information. Z.G., Lin Li, X.L., Jiaqian Wang, Y.Z., G.C., J.Y., L.C., M.H., M.L., X.H., Xuehan Zhuang, K.Q., G.Y. and G.G. performed sequencing and data analysis. Lin Li and K.H. performed the validation of variations. Y.O. performed experiments and data analysis, and wrote the manuscript. W.Z. performed *MIR548K* assays and analysed structural variation data. X.M., Lingyan Liu, W.Z., J.F., L.D., Z.Z. and Liying Ma performed *FAM135B* assays. Z.G., Lin Li and X.L. edited the manuscript. Lin Li and Jiaqian Wang performed the analysis of supplementary data. Ling Ma, J.Z., Longhai Luo, M.F., B.X., T.T., M.W., Z.L., D.L., Q.F. and P.C. provided supervision and support in the project. Y. L., Xiuqing Zhang, H.Y. and Jun Wang granted as well as supervised and supported this project.

Author Information Sequencing and array-CGH data have been deposited to the European Genome-phenome Archive (EGAS00001000709) and Gene Expression Omnibus (GSE54995). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Q.Z. (zhanqimin@pumc.edu.cn).

METHODS

Sample collection and preparation. This project was part of the International Cancer Genome Consortium (ICGC) (http://icgc.org/icgc/cgp/72/371/1001734). All the patients were from Chaoshan District of Guangdong Province, which is one of areas with highest prevalence of ESCC in China. Samples were collected in Cancer Hospital of Shantou University Medical College with documented informed consent through the institutional review board during 2007–11. Tumour and peripheral blood samples were collected from each patient who was undergoing surgical resection; none of the patients was treated by chemotherapy or radiotherapy before the operation. After being checked by the pathologist, the tumour tissues and regional lymph nodes were immediately frozen in liquid nitrogen and stored at −80 °C for further research. Meanwhile, partial tissues were used for haematoxylin and eosin staining to confirm the diagnosis, and analysis of pathological grades, metastasis and tumour cell content. All the tumour samples contained more than 80% tumour tissues free of necrosis. DNA was extracted from the frozen tissues and peripheral blood lymphocytes using either the QIAamp DNA Mini kit (Qiagen) or Wizard Genomic DNA Purification Kit (Promega). 200 ng genomic DNA was analysed on 1% agarose gel to ensure no sign of RNA contamination or degradation. Mass spectrometric fingerprint genotyping of 21 common SNPs was used to verify that both tumour DNA and normal DNA were derived from the same patient.

Library preparation and whole-genome sequencing. The qualified genomic DNA of tumour and matched peripheral blood from 17 ESCC patients were fragmented by an ultrasonicator Covaris E-210 (Covaris). By adjusting shearing parameters, DNA fragments were concentrated in 500-bp peaks with two libraries for each sample. These fragments were purified, end blunted, 'A' tailed, and adaptor ligated. 10 to 12 cycles of PCR were performed after size selection in the gel. The concentration of the libraries was quantified by a bioanalyser (Agilent Technologies) and real-time PCR method using ABI StepOne plus real-time PCR system (Life technologies). Paired end, 90-bp read-length sequencing was performed in the HiSeq 2000 sequencer according to manufacturer's instructions (Illumina).

Exome capture, library preparation and sequencing. The qualified genomic DNA from tumour and matched peripheral blood from 71 ESCC patients was fragmented by Covaris technology with resultant library fragments 200–300 bp, and then adapters were ligated to both ends of the fragments. Extracted DNA was then amplified by ligation-mediated PCR (LM-PCR), purified, and hybridized to the NimbleGenEZ 44M human exome array for enrichment, non-hybridized fragments were then washed out. Both non-captured and captured LM-PCR products were subjected to real-time PCR to estimate the magnitude of enrichment. Each captured library was then loaded on a Hiseq2000 platform, and we performed high-throughput sequencing for each captured library independently to ensure that each sample met the desired average fold coverage. Raw image files were processed by Illumina base calling Software 1.7 for base calling with default parameters and the sequences of each individual were generated as 90-bp paired-end reads.

CGH array processing and analysis. NimbleGen CGH platform: DNA labelling, hybridization, washing, array scanning and data analysis were carried out according to the NimbleGen CGH Arrays User's Guide and performed at CapitalBio Corporation. Briefly, pairs of genomic DNA (500 ng) were labelled with fluorescent dyes Cy3 (test samples) or Cy5 (reference DNA), samples were co-hybridized to Human CGH 3×720K Whole-Genome Tiling v3.0 Array (NimbleGen), with a median probe spacing of 2,509 bp. The arrays were scanned using MS200 scanner (NimbleGen) with 2 μm resolution, and fluorescent intensity data was extracted with NimbleScan 2.6 software. The hybridization controls (STC, Sample Tracking Controls) were used to confirm that the correct sample was hybridized to each array.

For each spot on the array, $\log_2$ ratios of the Cy3-labelled test sample versus Cy5 reference sample were computed. Before normalization and segmentation analysis, spatial correction was applied, which corrected position-dependent non-uniformity of signals across the array, specifically, locally weighted polynomial regression (LOESS) was used to adjust signal intensities based on X, Y feature position[29]. Normalization was then performed using the q-spline method[30], compensated for inherent differences in signal between the two dyes, followed by segmentation using the CNA calling algorithm segMNT. The segMNT algorithm identified copy number variation using a dynamic programming process that minimized the squared error relative to the segment means[27].

Agilent CGH platform: After DNA was labelled with fluorescent dyes Cy3 (test samples) or Cy5 (reference DNA), using the SureTag Complete DNA Labelling kit (Agilent), the individually labelled test and reference samples were then purified using Microcon YM-30 filters (Millipore). After that, the Cy3-labelled test DNA and Cy5-labelled reference DNA were mixed and combined with 2 × hybridization buffer, 10 × blocking agent (Agilent), and human Cot-1 DNA. The hybridization mixture was carefully dispensed to microarray chip (Agilent 1M) and assembled with Agilent SureHyb chamber. Then, the slide chamber was placed in the rotator rack in a hybridization oven set to 65 °C for 40 h with suitable rotation. After washing, all microarray slides were scanned on an Agilent Microarray Scanner G2565C with 2 μm resolution. Data were extracted from scanned images by the Feature Extraction software, version 11 (Agilent) and DNAcopy[26] were used to segmentation analysis.

Reads mapping and variation detection. After removing reads containing sequencing adapters and low-quality reads with more than five ambiguous bases, high-quality reads were aligned to the NCBI human reference genome (hg19) using BWA (v0.5.9)[21] with default parameters. Picard (v1.54) (http://picard.sourceforge.net/) was used to mark duplicates and followed by Genome Analysis Toolkit (v1.0.6076, GATK IndelRealigner)[23] to improve alignment accuracy.

Somatic point mutations were detected by VarScan2.2.5 (samtools (v0.1.18)[31] mpileup –Q 0 && VarScan2.2.5 somatic –min-coverage 10 –min-coverage-normal 10 –min-coverage-tumour 10 –min-var-freq 0.1 –min-avg-qual 0)[22]. Somatic indels were predicted with GATK SomaticIndelDetector with default parameters. All high-confident mutations were obtained using in-house pipeline coupled with visual inspection, then annotated with ANNOVAR (released 2 October 2011)[24].

Copy number alterations (CNAs) were detected by SegSeq[25] for WGS data. Then we combined WGS and a-CGH CNAs for further analysis. The resulting copy number segment statuses were defined as copy ratio ≥1.25 for gain and ≤0.75 for loss. To infer recurrently amplified or deleted genomic regions, we re-implemented the GISTIC algorithm[32] using copy numbers in 1-kb windows instead of SNP array probes as markers. G-scores were calculated for altered genomic regions based on the frequency and amplitude of amplification and deletion. A significant CNA region was defined as corresponding to a P-value threshold of 0.05 from permutation-derived null distribution, and peak regions for further analysis. And we compared ESCC CNA data to EAC, HNSCC and lung SQCC, which were downloaded from ICGC or TCGA data.

CREST[28] was used to predict somatic SVs breakpoints. To identify virus integration, unmapped reads pair were remapped to hg19 human reference genome and virus reference genomes. Paired reads with mapping quality ≥30, one end uniquely mapped to hg19 and the other end uniquely mapped to virus genome were clustered to determine the hg19-virus breakpoints. Paired reads with mapping quality ≥30 and uniquely mapped to virus were clustered to determine the free-virus.

Validation of ESCC. Some somatic indels were validated by PCR amplification and Sanger sequencing. The PCR primers for putative somatic variants were designed by primer 3[33] in silicon and initially used to amplify the source DNA from the tumours. PCR was performed on a Dual 96-well GeneAmp PCR System 9700 (Applied Biosystems), 20 ng template DNA from each sample was used per reaction. The products were sequenced by a 3730xl DNA Analyzer (Applied Biosystems). All sequences were analysed by the Sequencing Analysis Software Version 5.2 (Applied Biosystems). If the mutations were successfully confirmed in the tumours but not identified in the match normal DNA, the somatic statuses would be determined successful validation.

Validation of SNVs and some indels by mass spectrometry were performed with the MassArray platform of Sequenom by determining their genotypes in the tumours and matched blood. The genotyping assay and base-calling procedures were performed as previously described. We considered the following criteria as genotyping assay failure: if the Sequenom software was unable to design primers for PCR amplification and base extension at the primer design stage; or if the observed peak for a given assay was not significant enough for a confident call at the base-calling stage.

Somatic structure variation breakpoints were validated by PCR amplification and Sanger sequencing. The PCR primers for putative somatic variants were designed by primer 3[33] in silicon and initially used to amplify the source DNA in the tumours from 200 bp flank near SV breakpoint regions. PCR was performed on Dual 96-well GeneAmp PCR System 9700 (Applied Biosystems), 20 ng template DNA from each sample was used per reaction. The products were sequenced by 3730xl DNA Analyzer (Applied Biosystems). All sequences were analysed by the Sequencing Analysis Software Version 5.2 (Applied Biosystems). If the breakpoints were successfully confirmed in the tumours but not identified in the match normal DNA, the somatic status of the structure variation breakpoints would be determined successful validation.

Potential significantly mutated genes identification. Two statistical approaches were used to identify significantly mutated genes that were mutated in multiple tumour samples. We first identified significantly somatic mutations using MutsigCV method[34], and found *TP53*, *RB1* and *CDKN2A* as significant mutated genes with $q < 0.1$. Meanwhile, we carried out analysis focusing on the mutations in our cohort, which have been previously reported two or more times in COSMIC territory (http://www.sanger.ac.uk/genetics/CGP/cosmic/) and obtained 15 candidate genes. We next used the method of ref. 33 to analyse these 15 genes and identified 8 significantly mutated genes ($q < 0.1$).

Statistical analysis for clinical pathological and genetic data. First, we divided each clinical feature into two statuses (status 0 and status 1) defined below. Gender: status 0, patients were female; status 1, patients were male. Pathological grade:

status 0 represents G1 or G2, status 1 represents G3 or G4. N: status 0, degree of spread to regional lymph nodes was N0; status 1, degree of spread to regional lymph nodes was N1, N2 or N3. Dead/survival: status 0, patients were alive during the last follow-up; status 1, patients were dead during the last follow-up. Smoking: status 0, non-smoker; status 1, former and current smoker. Drinking: status 0, patients who never drank alcohol; status 1, former and current drinkers. Age: status 0, patients were younger than 59 years old; status 1, patients were older than 59 years old. Stage: status 0, stage I or II; status 1, stage III. For each classification of clinical features, significance of the correlation between mutated events and clinical features were calculated using the one-sided Fisher's exact test. The overall survival was estimated using the Kaplan–Meier method and the difference was tested using the log-rank test. A Cox proportional hazard model was used to analyse the hazards, while adjusting for clinical variables significant at $P < 0.1$ level in univariate analysis, as implemented in the R package 'survival' (http://cran.r-project.org/web/packages/survival/). All statistical analyses were performed using the computing environment R (R Development Core Team, 2005).

In the study, for both the Fisher's test and survival analysis, correction for multiple testing was done by using the Benjamini–Hochberg false discovery rate procedure.

Cell culture. Oesophageal squamous cell carcinoma (ESCC) cell lines such as KYSE30, KYSE70, KYSE180, KYSE410, KYSE450, KYSE140, COLO680N and KYSE510 were cultured in RPMI 1640 (Gibco) with 10% fetal bovine serum (FBS). ESCC cell line KYSE150 was cultured in a 1:1 mixture of RPMI 1640 and F12 (Gibco) culture medium with 2% FBS. Immortalized oesophageal epithelium cell lines NE2 and NE3 were cultured in a 1:1 mixture of EpiLife and dKSFM (Gibco) as described[35]. The fetal oesophageal epithelium cell line H5E973 was cultured according to the provider's recommendations[36]. All of these cells were maintained at 37 °C with 5% CO_2. We have recently authenticated the source of cell lines and tested for mycoplasma contamination, and did not find cross-contaminated cell lines and mycoplasma contamination.

Oligonucleotide transfection. Specific siRNA targeting *FAM135B* was custom-designed and provided by Ribobio. pGV141-FAM135B wild-type and mutated plasmids were constructed by Genechem. Hsa-miR-548K mimics (micrON hsa-miR-548k mimic, miR10005882-1-5) and inhibitors (micrOFF hsa-miR-548k inhibitor, miR20005882-1-5) were designed and provided by Ribobio. 10×10^4 cells were seeded on 60-mm culture plates and transfected with oligonucleotides using lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Transfected cells were incubated at 37 °C with 5% CO_2 for 48 h.

Real-time PCR assay. The real-time PCR was performed using the Premix Ex Taq kit (Takara) and a 7300 real-time PCR system (Life Technologies) according to manufacturer's instructions. The mRNA expression level of *FAM135B* was normalized to the endogenous expression of *GAPDH*. Primers were provided by Invitrogen as described below: *FAM135B* forward, 5′-AGTGCTGACCTACATTGAC-3′; *FAM135B* reverse, 5′-GGCTTCCACCTACTACCA-3′. GAPDH forward, 5′-TCTCTGCTCCTCCTGTTC-3′, GAPDH reverse, 5′- GTTGACTCCGACCTTCAC-3′. Bulge-Loophsa-miR-548k qRT–PCR Primer Set (Ribobio, miRQ0005882-1-2) was used for the measurement of the relative quantity of hsa-miR-548k.

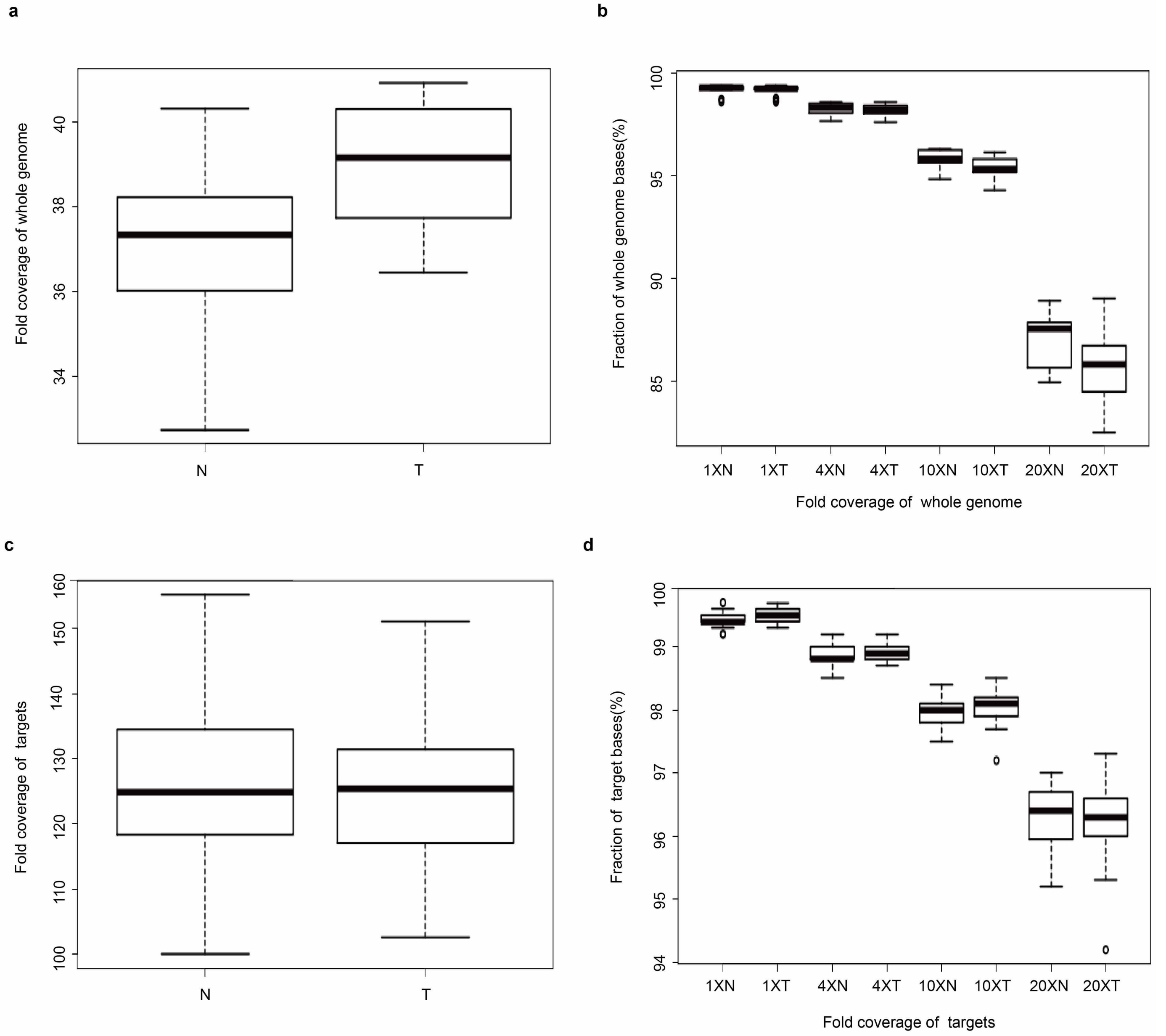
Proliferation assay. The proliferation ability of different cancer cells was monitored by using the xCELLigence Real-Time Cell Analyzer (RTCA)-MP system (Acea Biosciences/Roche Applied Science). This platform is able to measure cellular growth status in real time. 50 µl of culture medium was added in each well of E-Plate 96 (Roche Applied Science) to obtain equilibrium. Transfected cells were incubated in 60 mm culture plates for 24 h and then 2,000 cells in 100 µl of culture medium were seeded in E-Plate 96. E-Plate 96 was locked in RTCA-MP device at 37 °C with 5% CO_2. Measured changes in electrical impedance were present as cell index that directly reflects cellular proliferation on biocompatible microelectrode coated surfaces[37]. Cell index was read automatically every 15 min and the recorded curve was shown as cell index ± s.e.m.

Transwell migration/invasion assays. Migration and invasion assays were performed as described[38]. In brief, migration of cells was assayed in Transwell cell culture chambers with 6.5-mm diameter polycarbonate membrane filters containing 8 µm pore size (Neuro Probe). Briefly, 4×10^4 cells in 100 µl of serum-free medium were added to the upper chamber of the device, and the lower chamber was filled with 600 µl culture medium with 20% fetal bovine serum. After 10 h of incubation at 37 °C, the non-migration cells were removed from the upper surface of the membrane with a cotton swab. The filters were then fixed in methanol for 10 min, stained with Giemsa solution for 1 h, and counted. Five random microscopic fields (×100) were counted per well and the mean was determined. For the transwell invasion assay, the membrane of the upper chamber was pre-coated with 50 µl of a 2.5 mg ml^{-1} solution of matrigel (Falcon BD).

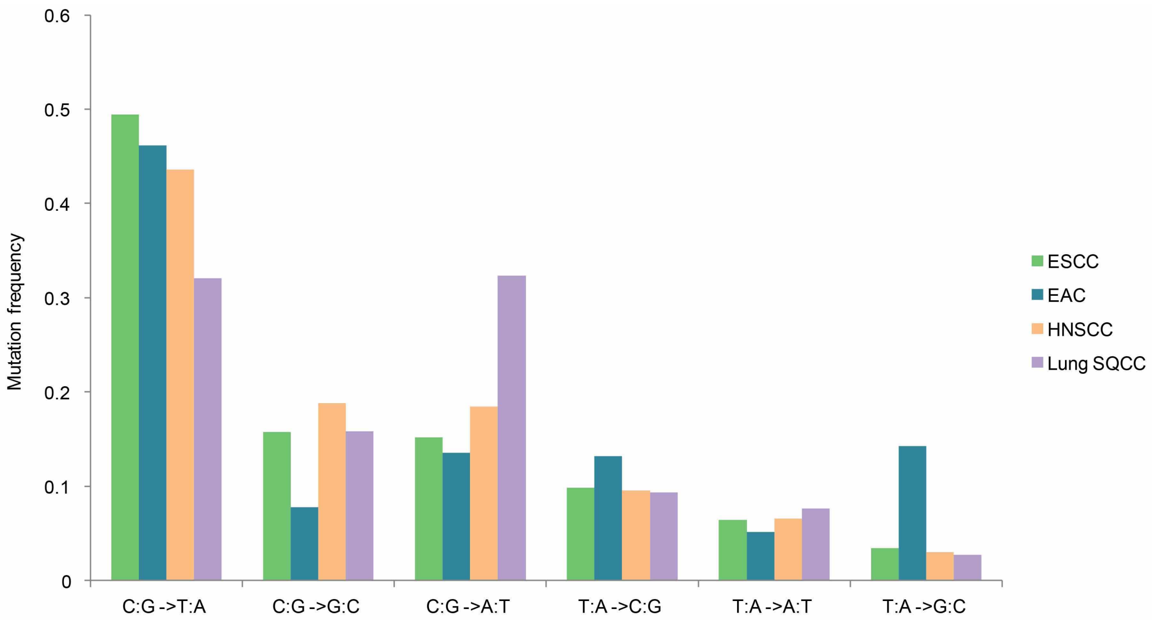
Colony formation assay. 2,000 transfected cells were seeded into 100-mm culture plates and incubated at 37 °C with 5% CO_2 for 10 days. Culture plates were performed in duplicates. After a wash with pre-cooled PBS, cultures were fixed with pre-cooled methanol for 20 min and stained with crystal violet for 15 min. Colonies were examined and automatically calculated by G:box (Syngene).

Statistics in the experiments. All experiments were performed in triplicate. Data were shown as mean value ± s.e.m. and statistically analysed by two-sided Student's *t*-test. $P < 0.05$ was considered as statistical significance.

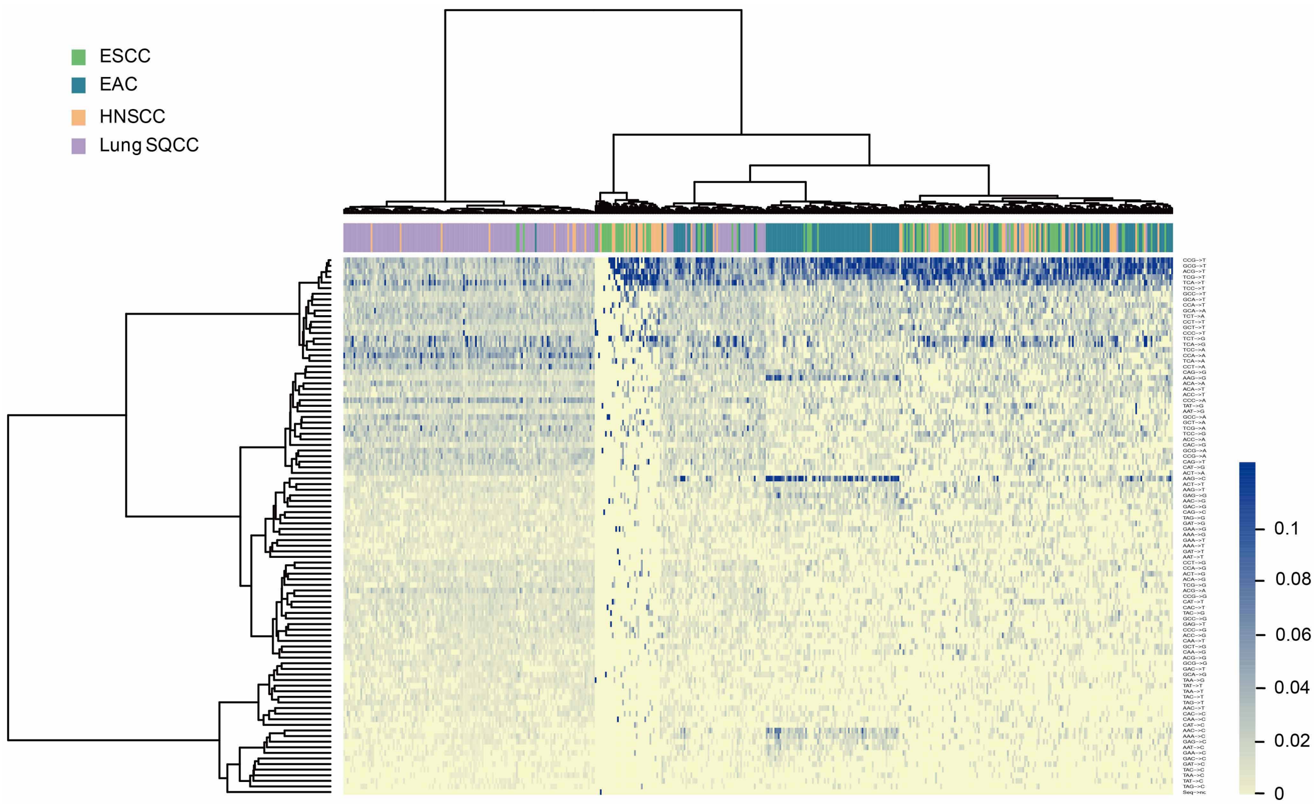
29. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31,** 265–273 (2003).
30. Workman, C. *et al.* A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* **3,** research0048 (2002).
31. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
32. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104,** 20007–20012 (2007).
33. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466,** 869–873 (2010).
34. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
35. Zhang, H. *et al.* Cytogenetic aberrations in immortalization of esophageal epithelial cells. *Cancer Genet. Cytogenet.* **165,** 25–35 (2006).
36. Shen, Z. Y. *et al.* Telomere and telomerase in the initial stage of immortalization of esophageal epithelial cell. *World J. Gastroenterol.* **8,** 357–362 (2002).
37. Wittchen, E. S. & Hartnett, M. E. The small GTPase Rap1 is a novel regulator of RPE cell barrier function. *Invest. Ophthalmol. Vis. Sci.* **52,** 7455–7463 (2011).
38. Ou, Y. *et al.* Migfilin protein promotes migration and invasion in human glioma through epidermal growth factor receptor-mediated phospholipase C-gamma and STAT3 protein signaling pathways. *J. Biol. Chem.* **287,** 32394–32405 (2012).



Extended Data Figure 1 | Fold coverage of whole genome and target regions in the sequenced normal and tumour samples in ESCC. **a**, The box plot depicts the distribution of mean coverage of all whole-genome sequencing samples. Lines in the two central boxes show the medians, and lines outside the two central boxes show the first and the third quartiles of the mean depths. **b**, The box plot depicts the distribution of fraction of whole-genome bases covered by at least 1 read, 4 reads, 10 reads and 20 reads across the 34 whole-genome sequencing samples. The lines in boxes show the medians and the lines outside the boxes show the first or third quartiles of fraction of whole-genome bases covered by reads. **c**, The box plot depicts the distribution of mean coverage of all whole-exome sequencing samples. **d**, The box plot depicts the distribution of fraction of targeted bases covered by at least 1 read, 4 reads, 10 reads and 20 reads across the 142 whole-exome sequencing samples. N, normal samples; T, tumour samples.

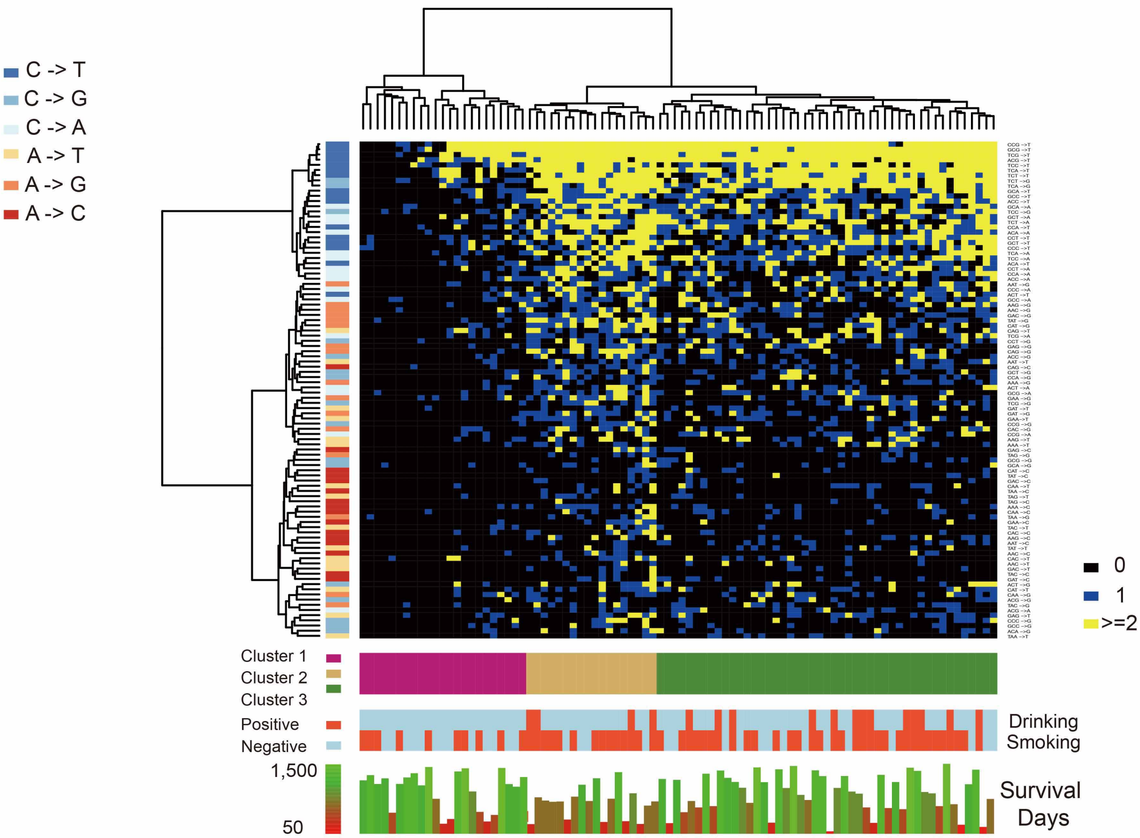


Extended Data Figure 2 | Spectrum of somatic point mutations identified in exome regions of ESCC, EAC, HNSCC and lung SQCC. Genomic data from 88 ESCC, 145 EAC, 74 HNSCC and 177 lung SQCC were analysed.

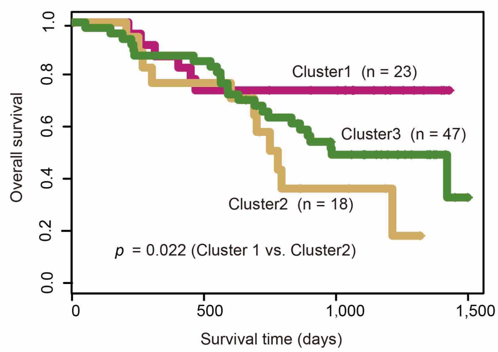


Extended Data Figure 3 | Hierarchical clustering of 484 samples in ESCC, EAC, HNSCC and lung SQCC according to their nucleotide context-specific exonic mutation rates. Top bar: cancer types of each sample. Genomic data from 88 ESCC, 145 EAC, 74 HNSCC and 177 lung SQCC were analysed.

a



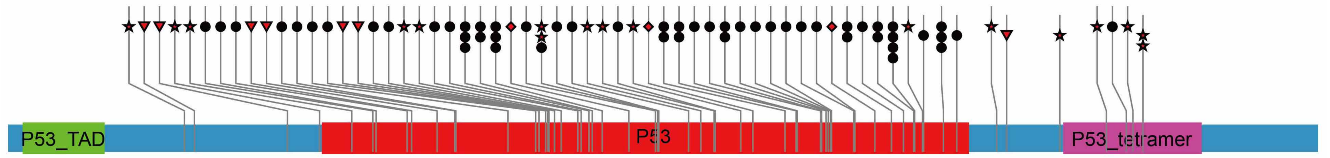
b



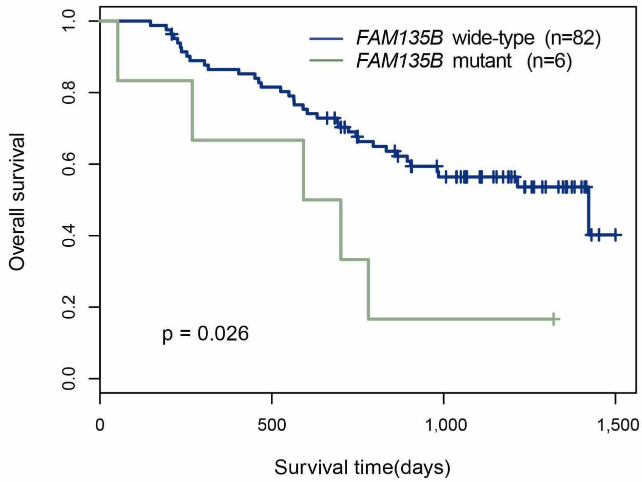
Univariate and multivariate analysis by the Cox proportional hazards model

Variable	Univariate analysis		Multivariate analysis	
	HR(95% CI)	*p* value	HR(95% CI)	*p* value
Age(<59 y vs. >=59 y)	0.731(0.281-1.901)	0.520		
Gender(male vs. female)	1.157(0.407-3.286)	0.784		
Pathological grade(G1 or G2 vs. G3 or G4)	4.023(1.279-12.680)	0.017	4.370(1.138-16.772)	0.032
Stage(I or II vs. III or IV)	1.715(0.661-4.454)	0.268		
Smoking(never smoked vs. former and current smoker)	1.863(0.655-5.299)	0.243		
Drinking(never drank vs. former and current drinking)	1.220(0.727-2.049)	0.010	4.545(1.097-18.828)	0.037
Cluster(cluster 1 vs. cluster 2)	3.045(1.118-8.294)	0.029	1.520(0.441-5.234)	0.507

Extended Data Figure 4 | Mutation spectrum analysis of 88 ESCCs. **a**, Context-specific mutation-based unsupervised clustering analysis in 88 ESCC cases. Heat map shows somatic mutation counts of specific mutation signatures in each case. Bottom bars: reported clusters, drinking and smoking status, and survival time. **b**, Top: Kaplan–Meier survival curve for three clusters of patients: pink line represents cluster 1 ($n = 23$); brown line represents cluster 2 ($n = 18$); and green line represents cluster 3 ($n = 47$). Cluster 1 patients had better prognosis as compared with patients of cluster 2 ($P = 0.022$, log-rank). Bottom: Cox proportional hazards model for cluster 1 and cluster 2 patients. $P < 0.05$ was considered statistically significant.



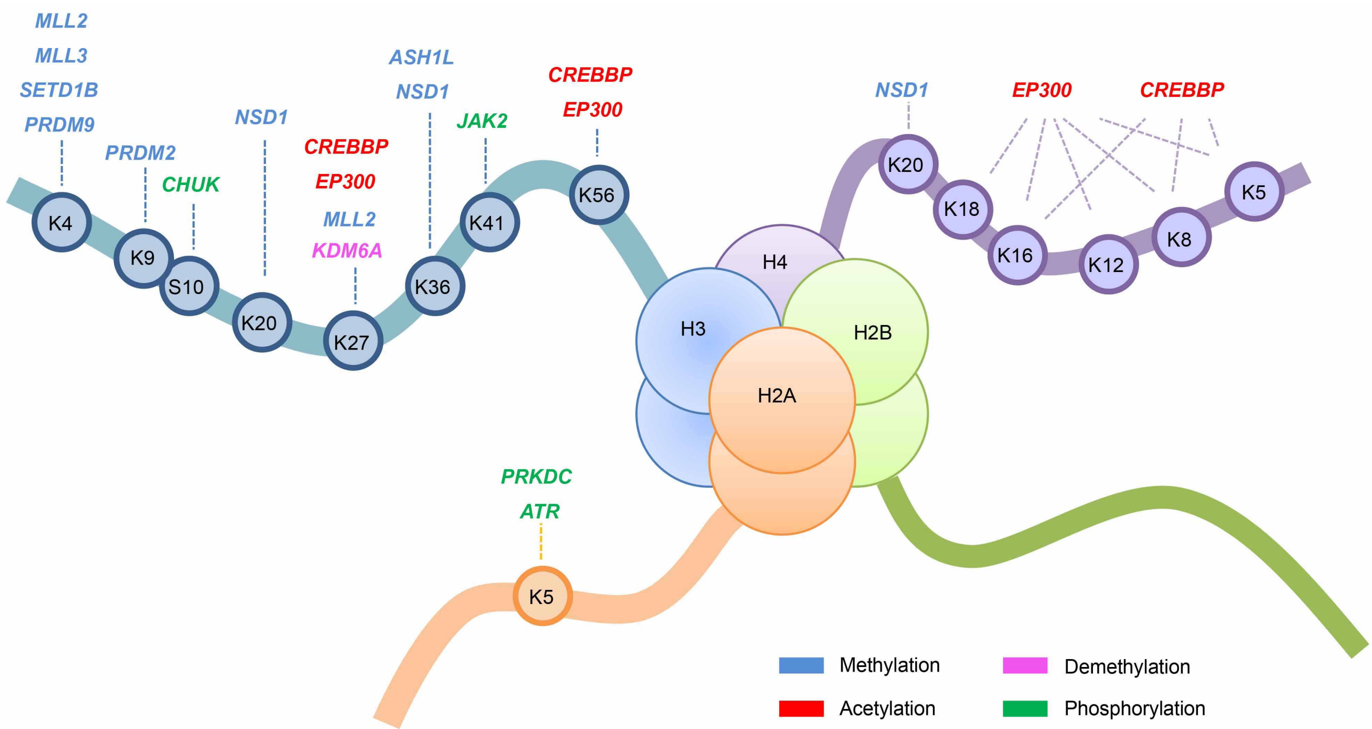
Extended Data Figure 5 | Somatic mutations in *TP53*. The types and relative positions of confirmed somatic mutations are shown in the transcript of *TP53*. Red stars, nonsense mutations ($n = 17$); bullets, missense mutations ($n = 53$); red triangles, indels ($n = 7$); and diamond, mutations at splice sites ($n = 3$). P53_TAD, p53 transactivation domain; P53, p53 DNA-binding domain; P53_tetramer, p53 tetramerization motif.



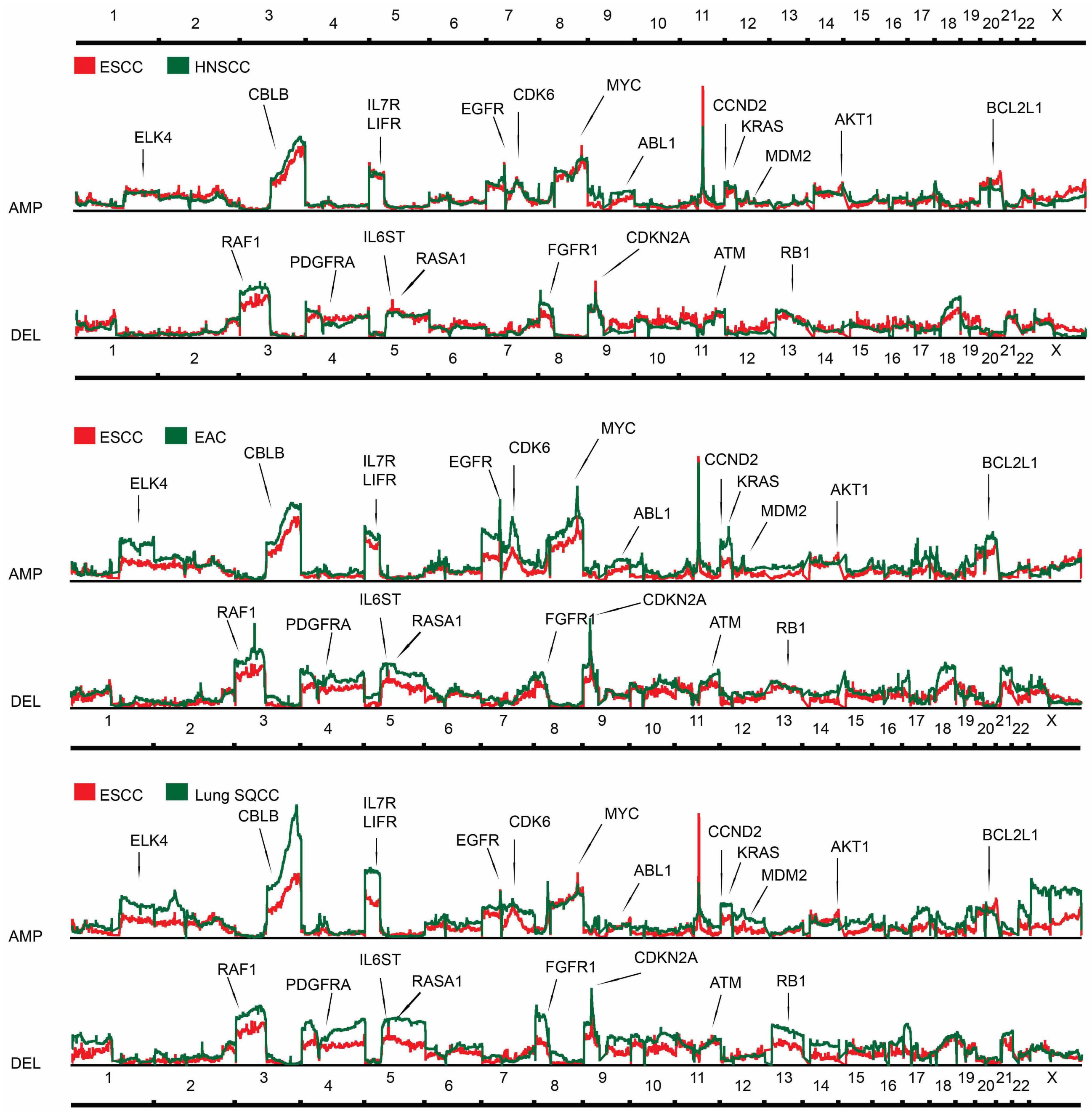
Univariate and multivariate analysis by the Cox proportional hazards model

Variable	Univariate analysis		Multivariate analysis	
	HR(95% CI)	*p* value	HR(95% CI)	*p* value
Age(<59 y vs. >=59 y)	1.380(0.740-2.573)	0.311		
Gender(male vs. female)	1.011(0.495-2.061)	0.978		
Pathological grade(G1 or G2 vs. G3 or G4)	3.266(1.357-7.860)	0.008	2.600(1.027-6.584)	0.044
Stage(I or II vs. III or IV)	2.710(1.454-5.049)	0.002	2.149(1.097-4.212)	0.025
Smoking(never smoked vs. former and current smoker)	1.688(0.844-3.375)	0.139		
Drinking(never drank vs. former and current drinking)	1.731(0.866-3.459)	0.121		
FAM135B(wide-type vs. mutant)	2.787(1.086-7.149)	0.033	2.214(0.816-6.012)	0.119

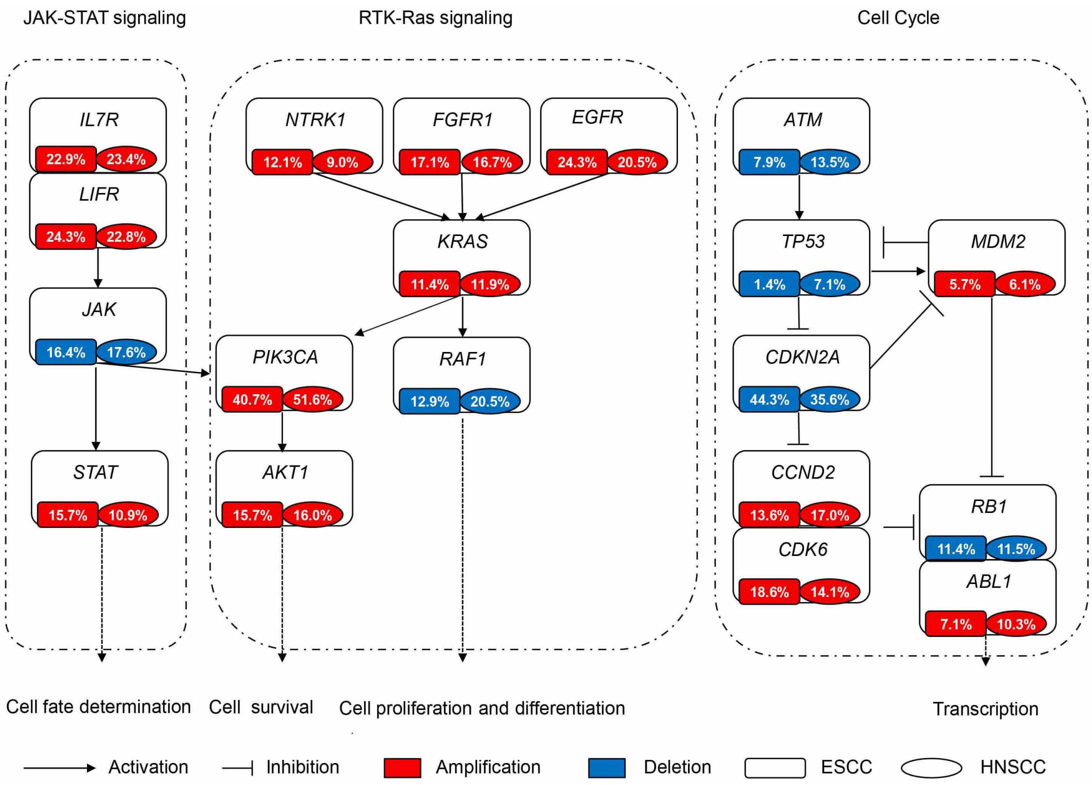
Extended Data Figure 6 | The relationship between survival time and mutations of *FAM135B* in ESCC patients. Top: Kaplan–Meier survival curve for wild-type and *FAM135B* mutant ($P = 0.026$, log-rank). Blue line, *FAM135B* wide type ($n = 82$); green line, *FAM135B* mutant ($n = 6$). Bottom: Cox proportional hazards model for wild-type and mutations of *FAM135B*. $P < 0.05$ was considered statistically significant.



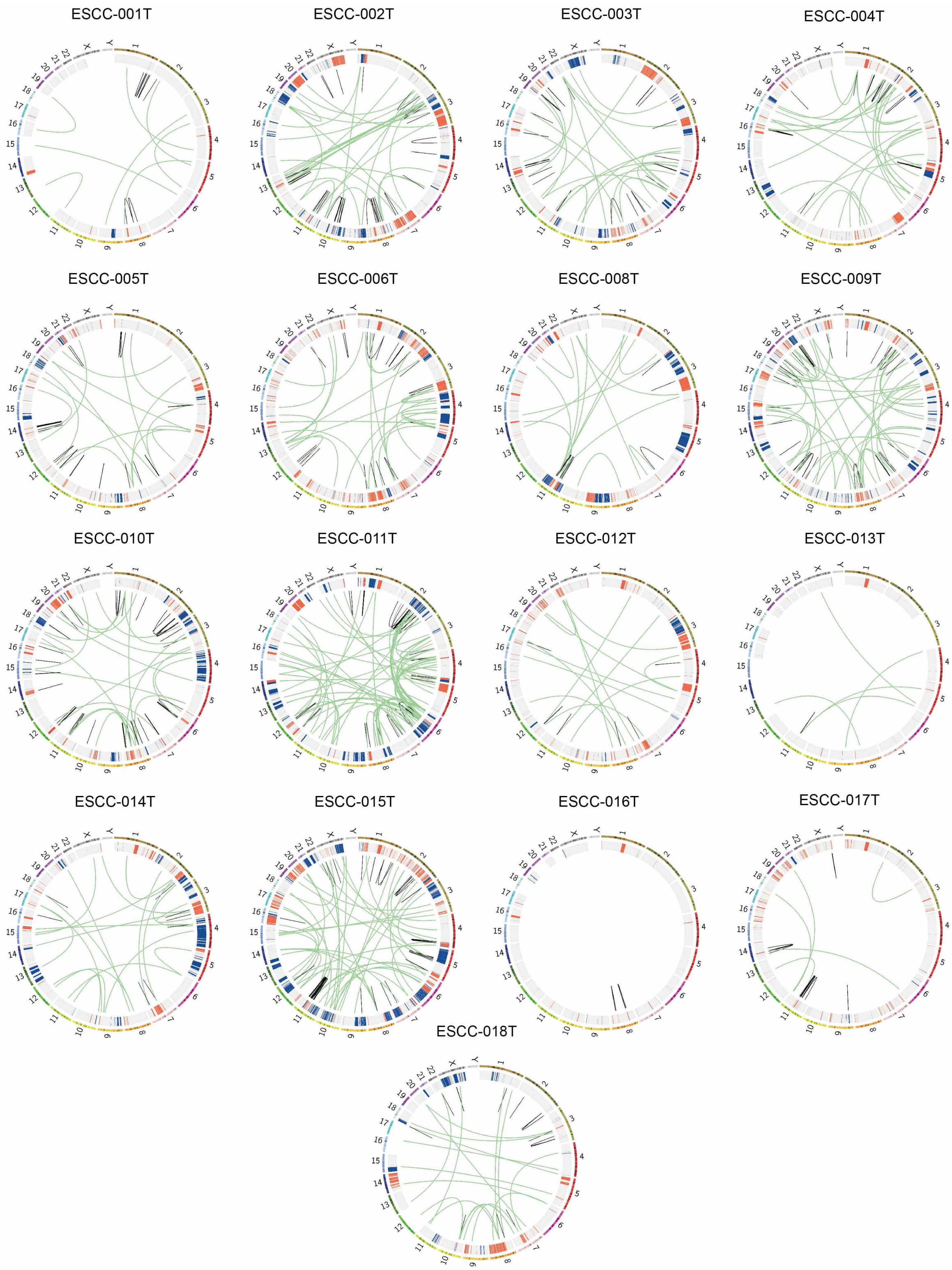
Extended Data Figure 7 | Histone-modifying genes recurrently mutated in 88 ESCCs. The sites for modification are marked in colour. Histone octamer with main methylation (blue), acetylation (red) and phosphorylation (green) genes on specific histone residues mutated in more than one sample are shown.



Extended Data Figure 8 | Comparative analysis of genomic copy number alterations among ESCC, EAC, HNSCC and lung SQCC. Genomic data from 140 ESCC, 70 EAC, 312 HNSCC, 663 lung SQCC were analysed. Figure shows the amplification (AMP) and deletion (DEL) for chromosomes 1–22 and X. High-frequency differences occurring in four cancer types are indicated in respective curves.



Extended Data Figure 9 | Copy number alterations with similar frequency identified between ESCC and HNSCC in JAK–STAT signalling, RTK–Ras signalling and cell cycle pathways. Frequency of copy number alterations are shown under genes. Rectangle, ESCC; ellipse, HNSCC.



Extended Data Figure 10 | Circos plot of intra- and inter-chromosomal translocations in all 17 WGS cases. Intra-chromosomal, green; inter-chromosomal, black.

doi:10.1038/nature13136

Cystathionine γ-lyase deficiency mediates neurodegeneration in Huntington's disease

Bindu D. Paul[1], Juan I. Sbodio[1], Risheng Xu[1,2], M. Scott Vandiver[1,2], Jiyoung Y. Cha[1], Adele M. Snowman[1] & Solomon H. Snyder[1,2,3]

Huntington's disease is an autosomal dominant disease associated with a mutation in the gene encoding huntingtin (Htt) leading to expanded polyglutamine repeats of mutant Htt (mHtt) that elicit oxidative stress, neurotoxicity, and motor and behavioural changes[1]. Huntington's disease is characterized by highly selective and profound damage to the corpus striatum, which regulates motor function. Striatal selectivity of Huntington's disease may reflect the striatally selective small G protein Rhes binding to mHtt and enhancing its neurotoxicity[2]. Specific molecular mechanisms by which mHtt elicits neurodegeneration have been hard to determine. Here we show a major depletion of cystathionine γ-lyase (CSE), the biosynthetic enzyme for cysteine, in Huntington's disease tissues, which may mediate Huntington's disease pathophysiology. The defect occurs at the transcriptional level and seems to reflect influences of mHtt on specificity protein 1, a transcriptional activator for CSE. Consistent with the notion of loss of CSE as a pathogenic mechanism, supplementation with cysteine reverses abnormalities in cultures of Huntington's disease tissues and in intact mouse models of Huntington's disease, suggesting therapeutic potential.

CSE is a principal generator of cysteine from cystathionine[3,4]. Cystathionine is formed by cystathionine β-synthase (CBS) by condensing homocysteine and serine. CSE, CBS and 3-mercaptopyruvate sulphurtransferase use cysteine to generate the major gasotransmitter hydrogen sulphide (H_2S)[3–5]. It was previously believed that CSE is restricted to peripheral tissues, whereas CBS is the principal generator of H_2S in the brain[6]. We detected substantial CSE in brain lysates, implying a role for the enzyme in the brain (Fig. 1a, b). In characterizing CSE-deleted mice[7] we noted abnormal hindlimb clasping and clenching reminiscent of mouse models of Huntington's disease, which prompted an exploration of CSE in Huntington's disease (Fig. 1c). In a striatal cell line Huntington's disease model containing 111 glutamine repeats, ST*Hdh*$^{Q111/Q111}$(Q111), CSE protein levels were markedly decreased compared with control ST*Hdh*$^{Q7/Q7}$(Q7) cells harbouring seven glutamine repeats (Fig. 1d, e). Brains of R6/2 mice with Huntington's disease[8] also showed decreased CSE levels in the striatum, cortex, hippocampus, hypothalamus and brainstem, but not in the cerebellum (Fig. 1f, g). To explore the generality of findings with R6/2 mice, we examined the Q175 mouse model of Huntington's disease[9]. Like the R6/2 mice, striata of the Q175 mice had diminished CSE levels (Fig. 1h, i). In brains of humans with Huntington's disease, CSE levels were profoundly reduced in the striatum (Fig. 1j–m), moderately diminished in the cerebral cortex and not altered in the cerebellum (Fig. 1l), paralleling the relative susceptibility of these brain regions to damage by Huntington's disease. Clinical staging revealed a progressive depletion of CSE with increasing severity of the disease (Fig. 1l, m). The depletion of CSE in Huntington's disease is not secondary to neurodegeneration. Thus, CSE levels were unaltered in post-mortem brain samples of patients with other neurodegenerative diseases: advanced amyotrophic lateral sclerosis, multiple sclerosis and spinocerebellar ataxia (Extended Data Fig. 1). In liver and pancreatic lysates of R6/2 mice, CSE levels were diminished to a similar extent to that in brain (Fig. 1n, o). CSE generates both cysteine and H_2S (Fig. 2a). In striatal Q111 cells we observed a diminished formation of both H_2S (Fig. 2b) and cysteine (Fig. 2c). Cysteine-free diets are lethal to CSE-deleted mice[10,11], and CSE deficiency of Q111 cells was associated with cytotoxicity in the absence of cysteine (Fig. 2d). On cysteine-free medium, 80% of Q111 cells die with only modest cell loss in Q7 cells. Supplementation with L-cysteine reversed the lethality (Fig. 2d), as did overexpression of CSE (Fig. 2e). Thus, the CSE depletion in Huntington's disease cells renders them hypersensitive to lethality associated with cysteine deprivation. Cysteine levels and H_2S generation were diminished in striata of R6/2 mice (Fig. 2f, g), and CSE activity was reduced in striata of Q175 mice (Fig. 2h).

How does Huntington's disease elicit CSE depletion? The loss of cytosolic CSE could reflect the translocation of CSE to insoluble aggregates, which occurs in Huntington's disease[12,13]. However, in Q111 cells CSE was depleted to similar extents in supernatant and particulate fractions (Fig. 3a). CSE is probably not depleted by proteasomal enzymes, because the proteasomal inhibitor MG132 failed to elevate the very low levels of CSE in Q111 cells (Fig. 3b).

Might depletion of CSE occur at a transcriptional level? Consistent with this possibility, analysis by PCR with reverse transcription (RT–PCR) revealed a marked decrease in mRNA levels for CSE in Q111 cells (Fig. 3c, d). Moreover, reporter gene assays revealed substantially reduced CSE promoter activity in Q111 cells (Fig. 3e). Thus, CSE transcription in Huntington's disease is evidently impaired at the level of transcriptional regulation. mHtt can elicit toxicity by disrupting transcriptional processes[14,15]: it has been reported[16] that mHtt binds to and inhibits specificity protein 1 (Sp1), a known transcription factor for CSE[17,18]. Overexpression of both Sp1 and its co-activator transcription initiation factor TFIID subunit 4 (TAF4) rescued the diminished messenger RNA and protein levels of CSE in Q111 cells (Fig. 3f–h) and improved growth in cysteine-free growth medium (Fig. 3i). Thus, CSE depletion in Huntington's disease seems to reflect an inhibition of Sp1 by mHtt, leading to decreased CSE transcription.

Huntington's disease is associated with heightened oxidative stress and mitochondrial dysfunction[19], and CSE has a function in mitochondrial homeostasis[20]. To assess whether CSE depletion elicits oxidative and mitochondrial stress, we treated wild-type and *Cse*$^{-/-}$ mice with 3-nitropropionic acid, a mitochondrial toxin that inhibits the complex II enzyme succinate dehydrogenase[21]. Treatment with 3-nitropropionic acid causes mitochondrial dysfunction and oxidative stress, thereby damaging striatal tissue and mimicking Huntington's disease[22]. *Cse*$^{-/-}$ mice were highly susceptible to 3-nitropropionic acid, showing elevated levels of protein carbonylation and nitration in the striatum and cortex (Extended Data Fig. 2). In addition, striatal Q111 cells showed greater susceptibility to hydrogen peroxide stress, and overexpressing CSE increased their resistance to H_2O_2 (Fig. 4a). If the pathophysiology of Huntington's disease reflects, at least in part, CSE depletion, we might predict neurobehavioural aberrations in CSE deleted mice. *Cse*$^{-/-}$ mice showed impaired rotarod performance (Fig. 4b). *Cse*$^{+/-}$ heterozygous mice also showed impaired motor functions, although not as severe as in the homozygous *Cse*$^{-/-}$ mice (Fig. 4b).

[1]The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [2]Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [3]Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

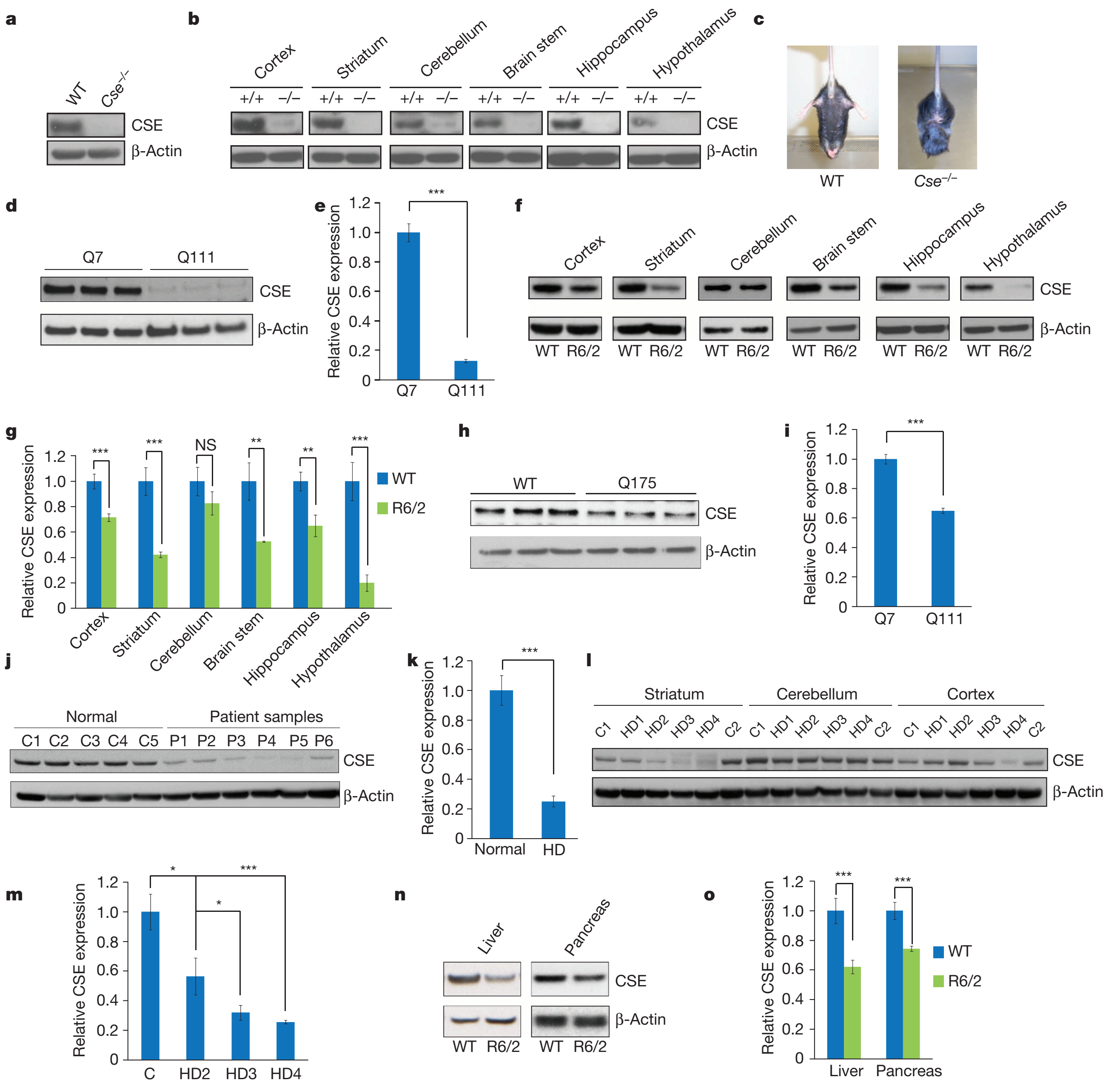


Figure 1 | CSE is expressed in the brain and is depleted in Huntington's disease. **a**, CSE expression is detectable in whole-brain lysates. WT, wild type. **b**, CSE is expressed in different regions of the brain. **c**, $Cse^{-/-}$ mice show a limb-clasping phenotype. **d**, Striatal $STHdh^{Q111/Q111}$ have decreased expression of CSE. **e**, Quantification of **d**. $n = 3$ (means ± s.e.m.); $***P < 0.001$. **f**, Expression of CSE is decreased significantly in all brain regions of 13-week-old R6/2 mice analysed except in the cerebellum. **g**, Relative quantification of **f**. $n = 6$ (means ± s.e.m.) for cortex, striatum and cerebellum; $n = 3$ (means ± s.e.m.) for other regions; $***P < 0.001$; $**P < 0.01$; NS, not significant. **h**, Expression of CSE is decreased in striata of Q175 mice. **i**, Quantification of **h**. $n = 3$ (means ± s.e.m.); $***P < 0.001$. **j**, Post-mortem striatal brain samples (P1–P6) from patients with Huntington's disease show a decrease in CSE expression. **k**, Relative quantification of **j**. $n = 5$ (means ± s.e.m.) for controls; $n = 6$ (means ± s.e.m.) for samples from patients with Huntington's disease (HD); $***P < 0.001$. **l**, CSE depletion is increased with the severity of the disease. C1 and C2 are controls; HD1–HD4 are samples from patients with Huntington's disease. **m**, Relative quantification of HD grades. $n = 3$ (means ± s.e.m.) for HD2; $n = 4$ (means ± s.e.m.) for normal, HD3 and HD4; $*P < 0.05$ (control versus HD2 and HD2 versus HD3); $***P < 0.001$ (control versus HD4). **n**, CSE expression is decreased in the liver and pancreas. **o**, Relative quantification of **n**. $n = 3$ (means ± s.e.m.); $***P < 0.001$.

By what molecular mechanisms might CSE exert cytoprotection? CSE is a major generator of cysteine, which is a building block for proteins and the antioxidant glutathione. In addition, cysteine is a precursor of H_2S, which signals by attaching a thiol group to target proteins, a process termed sulphydration; this process typically activates enzymes, some of which are involved in cytoprotection[3,23,24]. For instance, H_2S sulphydrates parkin to stimulate its catalytic activity, which provides striatal neuroprotection relevant to Parkinson's disease[25]. H_2S also sulphydrates Keap1 (ref. 26), a repressor of the transcription factor Nrf2, which regulates several enzymes of the antioxidant defence pathway. In addition, H_2S influences several cytoprotective pathways[27]. It is unclear whether the pathophysiological influences of CSE depletion in Huntington's disease reflect its role in generating cysteine or H_2S. It is conceivable that treatment with H_2S donors will be beneficial in treating Huntington's disease.

The ability of CSE and cysteine to reverse oxidative stress and lethality in Huntington's disease cells suggests that cysteine supplementation might be beneficial in treating Huntington's disease. Cysteine deficiency

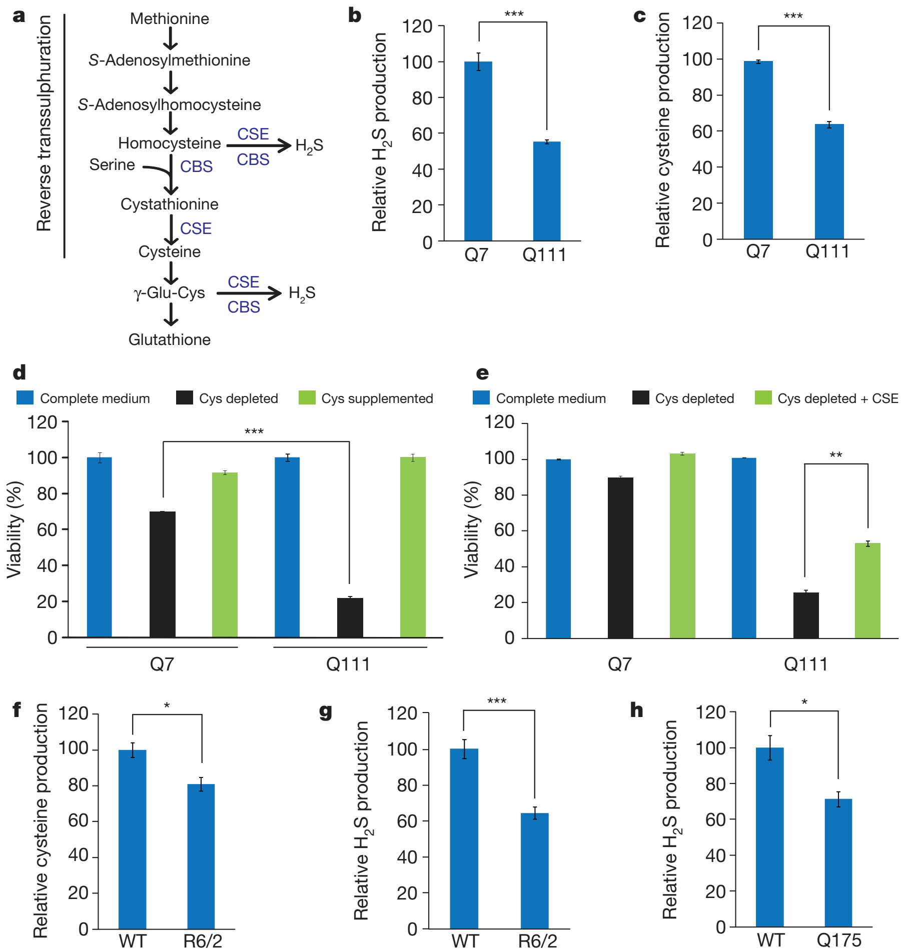


Figure 2 | Decreased CSE activity and growth in striatal Q111 cells. **a**, The reverse transsulphuration pathway leading to the production of cysteine from methionine. CSE produces cysteine from cystathionine generated by CBS. Cysteine and homocysteine are substrates for the production of H_2S. **b**, Decreased H_2S production in Q111 cells in comparison with Q7 cells. $n = 3$ (means ± s.e.m.); $***P < 0.001$. **c**, Decreased cysteine synthesis in Q111 cells. $n = 3$ (means ± s.e.m.); $***P < 0.001$. **d**, Impaired growth of Q111 cells in cysteine-free medium as monitored by the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2*H*-tetrazolium bromide (MTT) assay. Q111 cells underwent cell death, which was rescued by supplementation with L-cysteine. $n = 3$ (means ± s.e.m.); $***P < 0.001$. **e**, Growth retardation of Q111 cells in cysteine-free medium is rescued by the transfection of CSE construct as assessed by the MTT assay. $**P < 0.01$. **f**, Decreased cysteine levels in striata of R6/2 mice. $n = 3$ (means ± s.e.m.); $*P < 0.05$. **g**, Decreased synthesis of H_2S in striata of R6/2 mice. $n = 3$ (means ± s.e.m.); $***P < 0.001$. **h**, Decreased synthesis of H_2S in striata of Q175 mice. $n = 3$ (means ± s.e.m.); $*P < 0.05$.

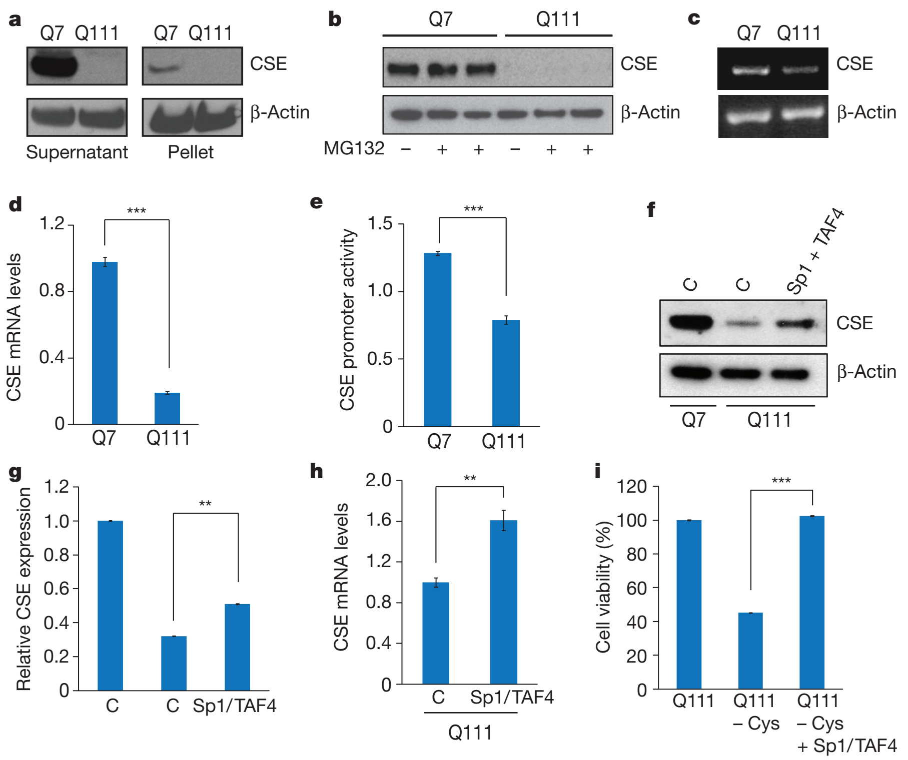


Figure 3 | CSE is depleted at the transcriptional level in Huntington's disease. **a**, CSE is not sequestered in the insoluble pellet fraction by mutant huntingtin. **b**, CSE is not differentially degraded in Q7 and Q111 cells after treatment with the proteasome inhibitor MG132. **c**, CSE mRNA is decreased in Q111 cells, as revealed by RT–PCR using β-actin as the internal control. **d**, Reduction of CSE expression in Q111 cells verified by real-time quantitative PCR. $n = 3$ (means ± s.e.m.); $***P < 0.001$. **e**, CSE promoter activity is repressed in Q111 cells as revealed by luciferase assays using a CSE-luc reporter construct and an endogenous internal control. $n = 4$ (means ± s.e.m.); $***P < 0.001$. **f**, Overexpression of the transcription factor Sp1 and its co-activator TAF4 rescues CSE expression. Empty vector controls are denoted by C. **g**, Relative quantification of **f**. $n = 3$ (means ± s.e.m.); $**P < 0.01$. **h**, Overexpression of Sp1 and TAF4 elevates CSE transcript levels as revealed by quantitative PCR. $n = 3$ (means ± s.e.m.); $**P < 0.01$. **i**, Overexpression of Sp1/TAF4 rescues lethality of Q111 cells in cysteine-free media as measured by the MTT assay. $n = 3$ (means ± s.e.m.); $***P < 0.001$.

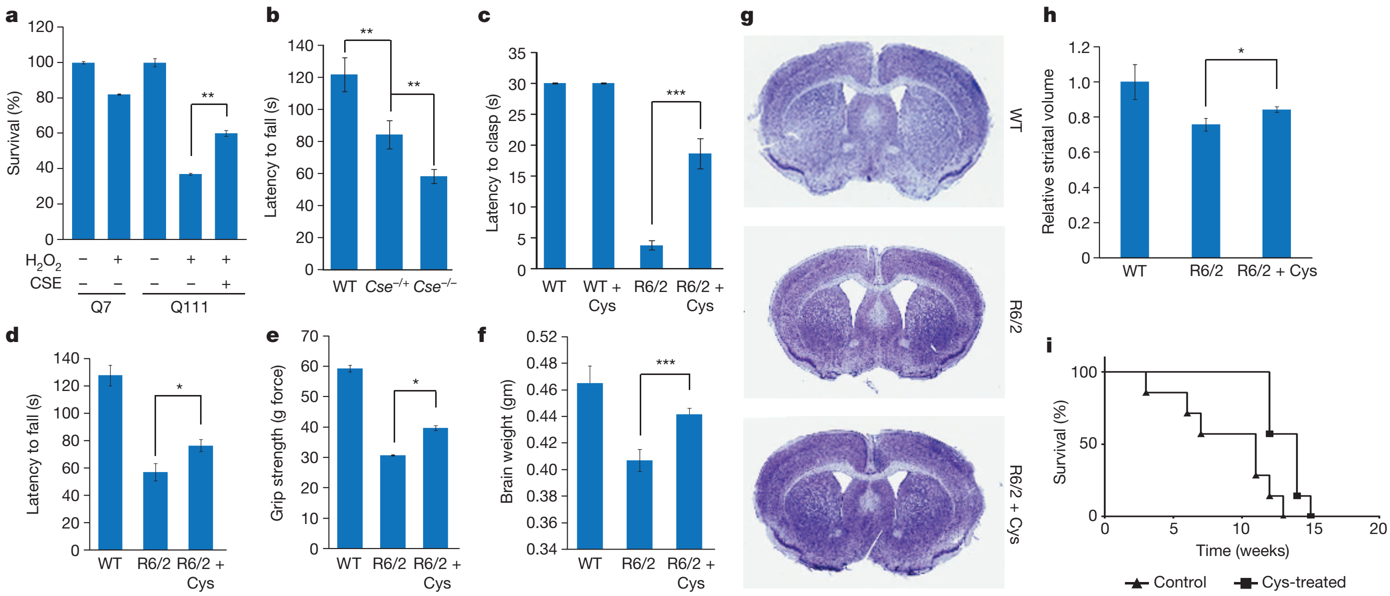


Figure 4 | CSE protects against oxidative stress, and cysteine supplementation delays neurodegeneration. **a**, Striatal Q111 cells are more vulnerable to oxidative stress induced by 0.1 mM H_2O_2 as monitored by MTT assays, effects that are rescued by CSE overexpression. $n = 3$ (means ± s.e.m.); $**P < 0.01$. **b**, $Cse^{-/-}$ mice have impaired motor function. WT, $Cse^{-/-}$ homozygous and $Cse^{+/-}$ heterozygous mice were placed on an accelerating rotarod, and latency to fall was recorded. Both the $Cse^{-/-}$ homozygous and $Cse^{+/-}$ heterozygous mice were impaired in their motor functions; the homozygous knockout mice showed the greatest deficits. $n = 5$ (means ± s.e.m.) for WT, $n = 4$ (means ± s.e.m.) for $Cse^{+/-}$ heterozygous and $n = 11$ for $Cse^{-/-}$ homozygous; $***P < 0.001$ (WT versus $Cse^{-/-}$); $**P < 0.01$ (WT versus $Cse^{+/-}$ and $Cse^{+/-}$ versus $Cse^{-/-}$). **c**, Cysteine supplementation delays motor symptoms in R6/2 mice. Mice were placed on regular diet or a cysteine-supplemented diet along with 20 mM *N*-acetylcysteine in the drinking water, and clasping phenotype was monitored. $n = 8$ (means ± s.e.m.); $***P < 0.001$. See also Supplementary Videos 1–3. **d**, Cysteine supplementation improves performance on an accelerating rotarod in R6/2 mice. $n = 8$ (means ± s.e.m.) for WT; $n = 6$ (means ± s.e.m.) for R6/2; $n = 7$ (means ± s.e.m.) for R6/2 + cysteine; $*P < 0.05$. **e**, Grip strength is improved in R6/2 mice placed on a cysteine-supplemented diet. $*P < 0.05$. **f**, Decrease in brain masses of R6/2 mice is ameliorated by cysteine treatment. **g**, Striatal atrophy is decreased in R6/2 mice treated with cysteine as assessed by Nissl staining of coronal sections of the brain. **h**, Striatal volume of R6/2 mice on a cysteine supplemented diet is larger than in the untreated R6/2 mice. $n = 3$ (means ± s.e.m.) for WT, $n = 4$ (means ± s.e.m.) for untreated R6/2 mice and R6/2 mice treated with cysteine. $*P < 0.05$. **i**, Cysteine supplementation prolongs survival in R6/2 mice (Kaplan–Meier analysis). R6/2 mice ($n = 7$ per group) were treated as in **f**, and survival in weeks was monitored. Statistical analysis was conducted with the log-rank (Mantel–Cox) test; $P = 0.004$.

has been implicated in oxidative stress and ageing[28]. Cystamine, the decarboxylated derivative of cystine (the disulphide of cysteine) confers neuroprotection in mouse models of Huntington's disease and increases cysteine levels in the brain[29]. Treating R6/2 mice with *N*-acetylcysteine (a stable precursor of cysteine) in their drinking water along with a cysteine-enriched diet delayed the onset of motor abnormalities such as hindlimb clasping (Fig. 4c and Supplementary Videos 1–3), rotarod performance (Fig. 4d) and grip strength (Fig. 4e). Decreases in brain weight and striatal were partly reversed by treatment with cysteine (Fig. 4f–h), and survival was enhanced (Fig. 4i).

Although many molecular targets of mHtt have been described, specific mechanisms that underlie Huntington's disease neurotoxicity have been hard to determine. It has been reported[16] that binding of mHtt to the transcriptional activator Sp1 mediates neurotoxicity by diminishing Sp1 activity. We propose that inhibition by mHtt of Sp1's transcriptional activation of CSE accounts for CSE depletion in Huntington's disease and associated neurological deficits. This CSE model of Huntington's disease is supported by the therapeutic effects of cysteine and *N*-acetylcysteine in mice with Huntington's disease. Our findings imply that *N*-acetylcysteine supplementation may be beneficial in treating diseases involving impaired reverse transsulphuration and oxidative stress.

METHODS SUMMARY

Assays of H_2S and cysteine were conducted as described previously, with minor modifications[7,20]. For the H_2S assay, tissue and cells were lysed in ice-cold 50 mM potassium phosphate buffer pH 6.8, containing 0.1% Triton X-100. Total protein (1 mg) was incubated with 10 mM L-cysteine, 15 μM pyridoxal 5-phosphate in tightly sealed vials flushed with nitrogen, and the reaction was incubated for 4 h at 37 °C in a total volume of 250 μl. Next, 125 μl of 1% zinc acetate and 2.5 μl of 10 M NaOH were added to the vials and shaken at 25 °C for 1 h to ensure trapping of H_2S released from the mixture. This was followed by the addition of 0.5 ml of distilled water and 0.1 ml of 20 mM *N*,*N*-dimethyl-*p*-phenylenediamine sulphate in 7.2 M HCl, immediately followed by the addition of 30 mM $FeCl_3 \cdot 6H_2O$ in 1.2 M HCl. The absorbance of the resulting product was measured at 670 nm. The Sp1 and TAF4 plasmid constructs were gifts from D. Krainc and harboured the respective genes under the CMV promoter, with a Myc tag and a haemagglutinin tag, respectively, for expression in mammalian cells.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May 2013; accepted 10 February 2014.
Published online 26 March 2014.

1. Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72,** 971–983 (1993).
2. Subramaniam, S. & Snyder, S. H. Huntington's disease is a disorder of the corpus striatum: focus on Rhes (Ras homologue enriched in the striatum). *Neuropharmacology* **60,** 1187–1192 (2011).
3. Paul, B. D. & Snyder, S. H. H_2S signalling through protein sulfhydration and beyond. *Nature Rev. Mol. Cell Biol.* **13,** 499–507 (2012).
4. Wang, R. Physiological implications of hydrogen sulfide: a whiff exploration that blossomed. *Physiol. Rev.* **92,** 791–896 (2012).
5. Kimura, H. Hydrogen sulfide: its production, release and functions. *Amino Acids* **41,** 113–121 (2011).
6. Morikawa, T. *et al.* Hypoxic regulation of the cerebral microcirculation is mediated by a carbon monoxide-sensitive hydrogen sulfide pathway. *Proc. Natl Acad. Sci. USA* **109,** 1293–1298 (2012).
7. Yang, G. *et al.* H_2S as a physiologic vasorelaxant: hypertension in mice with deletion of cystathionine γ-lyase. *Science* **322,** 587–590 (2008).
8. Mangiarini, L. *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87,** 493–506 (1996).

9. Menalled, L. B. *et al.* Comprehensive behavioral and molecular characterization of a new knock-in mouse model of Huntington's disease: zQ175. *PLoS ONE* **7,** e49838 (2012).
10. Ishii, I. *et al.* Cystathionine γ-lyase-deficient mice require dietary cysteine to protect against acute lethal myopathy and oxidative injury. *J. Biol. Chem.* **285,** 26358–26368 (2010).
11. Mani, S., Yang, G. & Wang, R. A critical life-supporting role for cystathionine γ-lyase in the absence of dietary cysteine supply. *Free Radic. Biol. Med.* **50,** 1280–1287 (2011).
12. DiFiglia, M. *et al.* Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277,** 1990–1993 (1997).
13. Davies, S. W. *et al.* Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **90,** 537–548 (1997).
14. Sugars, K. L. & Rubinsztein, D. C. Transcriptional abnormalities in Huntington disease. *Trends Genet.* **19,** 233–238 (2003).
15. Zhai, W., Jeong, H., Cui, L., Krainc, D. & Tjian, R. *In vitro* analysis of huntingtin-mediated transcriptional repression reveals multiple transcription factor targets. *Cell* **123,** 1241–1253 (2005).
16. Dunah, A. W. *et al.* Sp1 and TAFII130 transcriptional activity is disrupted in early Huntington's disease. *Science* **296,** 2238–2243 (2002).
17. Ishii, I. *et al.* Murine cystathionine γ-lyase: complete cDNA and genomic sequences, promoter activity, tissue distribution and developmental expression. *Biochem. J.* **381,** 113–123 (2004).
18. Yang, G., Pei, Y., Teng, H., Cao, Q. & Wang, R. Specificity protein-1 as a critical regulator of human cystathionine γ-lyase in smooth muscle cells. *J. Biol. Chem.* **286,** 26450–26460 (2011).
19. Lin, M. T. & Beal, M. F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443,** 787–795 (2006).
20. Fu, M. *et al.* Hydrogen sulfide (H_2S) metabolism in mitochondria and its regulatory role in energy production. *Proc. Natl Acad. Sci. USA* **109,** 2943–2948 (2012).
21. Alston, T. A., Mela, L. & Bright, H. J. 3-Nitropropionate, the toxic substance of *Indigofera*, is a suicide inactivator of succinate dehydrogenase. *Proc. Natl Acad. Sci. USA* **74,** 3767–3771 (1977).
22. Brouillet, E., Jacquard, C., Bizat, N. & Blum, D. 3-Nitropropionic acid: a mitochondrial toxin to uncover physiopathological mechanisms underlying striatal degeneration in Huntington's disease. *J. Neurochem.* **95,** 1521–1540 (2005).
23. Mustafa, A. K. *et al.* H_2S signals through protein S-sulfhydration. *Sci. Signal.* **2,** ra72 (2009).
24. Sen, N. *et al.* Hydrogen sulfide-linked sulfhydration of NF-κB mediates its antiapoptotic actions. *Mol. Cell* **45,** 13–24 (2012).
25. Vandiver, M. S. *et al.* Sulfhydration mediates neuroprotective actions of parkin. *Nature. Commun.* **4,** 1626 (2013).
26. Yang, G. *et al.* Hydrogen sulfide protects against cellular senescence via S-sulfhydration of Keap1 and activation of Nrf2. *Antioxid. Redox Signal.* **18,** 1906–1919 (2013).
27. Szabo, C. Hydrogen sulphide and its therapeutic potential. *Nature Rev. Drug Discov.* **6,** 917–935 (2007).
28. Droge, W. Oxidative stress and ageing: is ageing a cysteine deficiency syndrome? *Phil. Trans. R. Soc. Lond. B* **360,** 2355–2372 (2005).
29. Fox, J. H. *et al.* Cystamine increases L-cysteine levels in Huntington's disease transgenic mouse brain and in a PC12 model of polyglutamine aggregation. *J. Neurochem.* **91,** 413–422 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. C. Troncoso and O. Pletnikova for providing the human post-mortem tissue samples; D. Krainc for the constructs CMV-SP1 and TAF4; M. MacDonald for the striatal Q7 and Q111 cell lines; and the Cure Huntington's Disease Initiative (CHDI) for the Q175 mice tissues. This work was supported by United States Public Health Service Grant MH18501 to S.H.S. and by the CHDI. M.S.V. and R.X. are supported by the National Institutes of Health Medical Scientist Training Program Award.

Author Contributions B.D.P. and S.H.S. designed the research. B.D.P., J.S., R.X., M.S.V. and J.C. conducted experiments. B.D.P., J.S. and R.X. analysed data. A.M.S. prepared plasmid constructs and provided technical assistance. B.D.P. and S.H.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H.S. (ssnyder@jhmi.edu).

METHODS

Plasmids, cell lines and chemicals. The plasmids encoding Sp1 and TAF4 were gifts from D. Krainc (Massachusetts General Hospital, Boston, Massachusetts). The plasmid encoding CSE was described previously. The striatal cell lines ST*Hdh*$^{Q7/Q7}$, expressing wild-type huntingtin, and ST*Hdh*$^{Q111/Q111}$, expressing mutant huntingtin with 111 glutamine repeats (referred to as Q7 and Q111 cells, respectively), were from M. MacDonald (Department of Neurology, Massachusetts General Hospital, Boston). Unless otherwise mentioned, all chemicals were from Sigma. Lipofectamine 2000 (Invitrogen) was used for all transfection studies.

Western blotting. Cells were lysed on ice for 15 min in buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% Triton X-100 and protease inhibitors, followed by centrifugation and recovery of the supernatant. The pellet fraction was resuspended in 2% SDS lysis buffer followed by sonication. Protein concentrations were estimated with the bicinchoninic acid assay (Thermo Scientific). Anti-CSE antibodies were described previously[7]. The human post-mortem samples were obtained from J. Troncoso (Brain Resource Center, Johns Hopkins University).

RNA isolation and RT–PCR. Total RNA was isolated from cells and tissues using Trizol reagent followed by the RNEasy Lipid Tissue kit (Qiagen). Conventional RT–PCR was performed with the Superscript One-Step RT–PCR kit (Invitrogen). Real-time quantitative PCR was performed with the TaqMan RNA-to-Ct 1-Step Kit and the Step One Plus instrument (Life Technologies).

H_2S production assay. The H_2S production assay was as described previously[7,20].

Cysteine production assay. Cysteine formed by CSE was quantified as described previously[7]. Ninhydrin reagent was prepared by dissolving 250 mg of ninhydrin in 6 ml of glacial acetic acid and 4 ml of 12 M HCl. Samples (200 µl) were mixed with 60 µl of 6% perchloric acid, and precipitated protein was removed by centrifugation. Next, 200 µl of the supernatant was mixed with 200 µl of glacial acetic acid and ninhydrin, then boiled for 10 min, cooled on ice and mixed with 95% ethanol to a final volume of 1 ml. The absorbance was measured at 560 nm.

Animals and treatments. Animals were housed on a 12 h light–dark schedule and received food and water *ad libitum*. The R6/2 (B6CBA-Tg(HDexon1)62Gpb/1J) transgenic mice were from the Jackson Laboratory. The *Cse*$^{-/-}$ mice were described previously[7]. All animals were treated in accordance with the recommendations of the National Institutes of Health and approved by the Johns Hopkins University Committee on Animal Care.

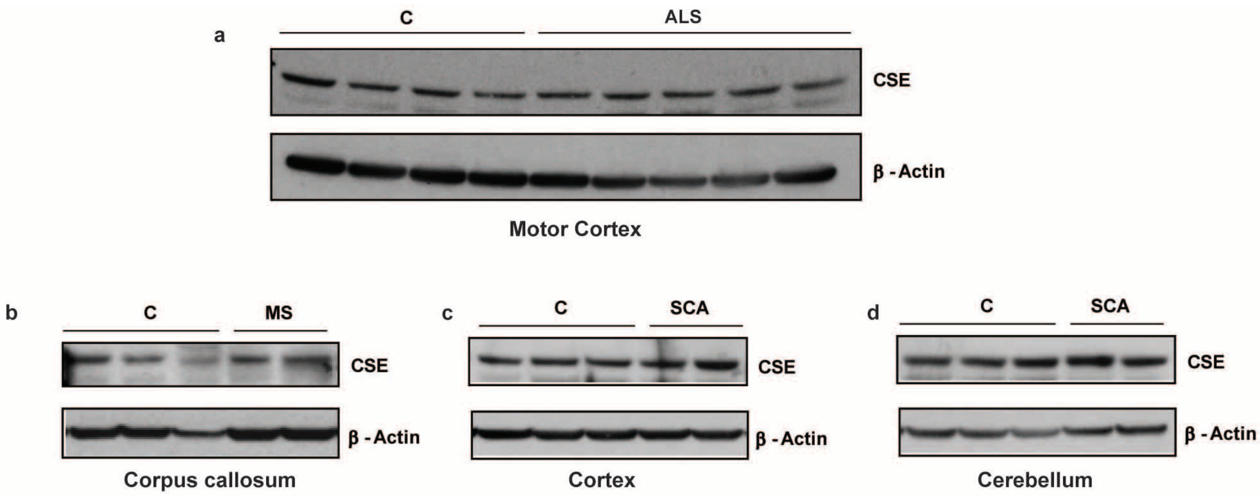
Cysteine treatment of R6/2 mice. All wild-type and R6/2 mice offspring derived from mice hemizygous for Tg(HDexon1)62Gpb (ovary transplant) were placed either on an AIN93 mineral diet with 0.3% cystine or on an AIN93 diet containing 0.8% cystine (Teklad Lab Animal Diets; Harlan Laboratories) starting at 4 weeks of age. In addition, the mice on diets containing high cysteine were also given 20 mM *N*-acetylcysteine in the drinking water. Survival and motor phenotypes of the mice were recorded.

Motor tests. The motor functions of the mice were tested with the hindlimb clasping function. Mice were suspended by their tail and the time taken to clasp was recorded. The accelerating rotarod was performed on a Rotamex V (Columbus Instruments) with speeds that varied from 4 to 40 r.p.m. for a maximum of 5 min with an acceleration interval of 30 s. Seven-week-old male R6/ 2 mice were used for the rotarod assay; 8-month-old male *Cse*$^{-/-}$ mice were used for the assay. The grip test was performed on 11-week-old male mice by placing the mouse over the baseplate by the tail such that its forelimbs grasped the metal grid. The tail was gently pulled backward until the grip was released. The force required to release the grip was recorded by the grip strength meter.

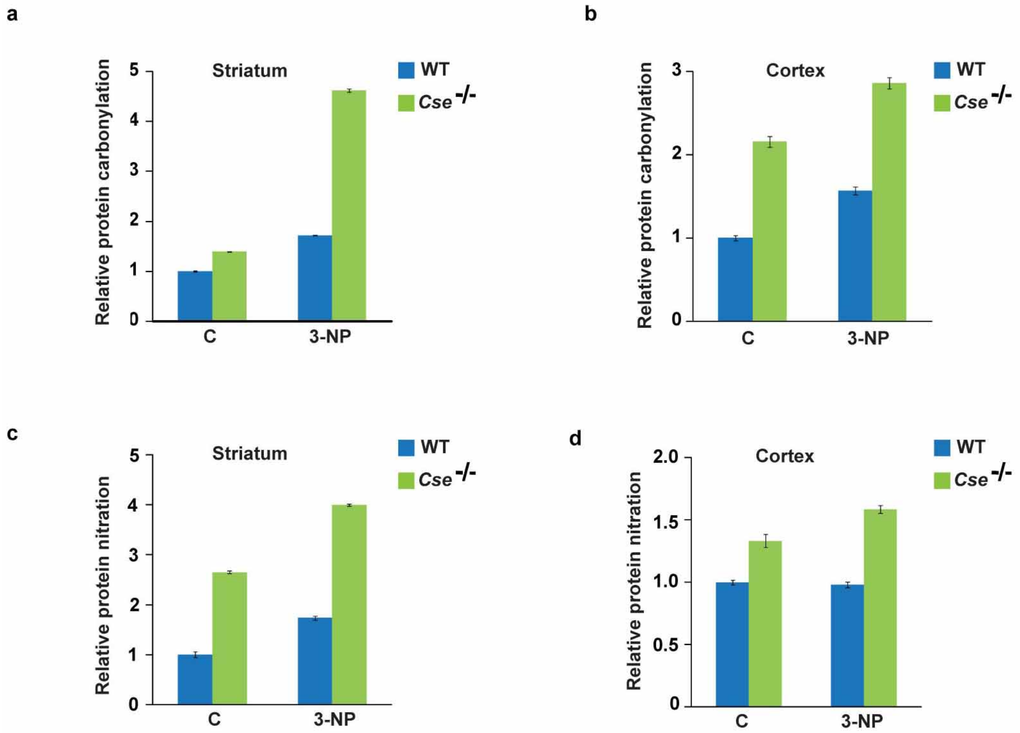
Statistical analyses. Results are presented as means ± s.e.m. for at least three independent experiments. The sample sizes used were based on the magnitude of changes and consistency expected. Statistical significance was reported as appropriate. In experiments involving animals, no exclusions were done. Sample sizes were chosen on the basis of the magnitude of changes expected. No randomization methods were used. In behavioural analyses, the experimenter conducting the test was blinded to the genotype or treatment of the animals under study. *P* values were calculated with Student's *t*-test.

Histology and quantification of relative striatal volume. Male mice at 12 weeks of age were perfused transcardially with 4% paraformaldehyde (PFA) in PBS and brains were stored overnight in 4% PFA in PBS. Next, brains were cryoprotected in 30% sucrose in PBS and sectioned with a freezing microtome at a section thickness of 30 µm. Sections were mounted on glass slides and allowed to dry at 25 °C for 24 h and then stained with Nissl stain. Striatal volumes were calculated using Cavalieri's principle[30] (volume $= s_1d_1 + s_2d_2 + \ldots + s_nd_n$, where s is surface area and d is the distance between two sections).

30. Cyr, M., Caron, M. G., Johnson, G. A. & Laakso, A. Magnetic resonance imaging at microscopic resolution reveals subtle morphological changes in a mouse model of dopaminergic hyperfunction. *Neuroimage* **26,** 83–90 (2005).



Extended Data Figure 1 | CSE expression is not altered in the brain in amyotrophic lateral sclerosis, multiple sclerosis and spinocerebellar ataxia. **a**, Western blots show that CSE expression in the motor cortex of samples from controls and patients with amyotrophic lateral sclerosis (ALS) showing substantial neurodegeneration in the motor cortex are similar. Extracts were prepared from the motor cortex and analysed for CSE expression using anti-CSE antibodies and β-actin as a loading control. **b**, Expression of CSE is not altered in the corpus callosum of patients with multiple sclerosis (MS), where multiple lesions, demyelination and decrease in oligodendrocytes was observed in the corpus callosum of the brain. **c**, **d**, Levels of CSE do not change in the cerebral cortex (**c**) or cerebellum (**d**) of patients with spinocerebellar ataxia (SCA). Neuropathological analysis of the brains of these patients revealed severe neuronal loss and gliosis in the cerebellum.



Extended Data Figure 2 | *Cse*$^{-/-}$ mice are more vulnerable to stress induced by 3-nitropropionic acid. Wild-type and *Cse*$^{-/-}$ male mice at 8 months of age were injected with a single dose of 3-nitropropionic acid (3-NP) (100 mg kg^{-1}), and lysates were prepared 24 h later from the striatum and cortex and analysed for oxidative stress. **a**, **b**, Striata (**a**) and cortex (**b**) of *Cse*$^{-/-}$ mice show elevated protein oxidation as measured by protein carbonylation, which is more pronounced in the striatum. $n = 3$ (means $\pm$ s.e.m.). **c**, **d**, *Cse*$^{-/-}$ mice also show augmented levels of protein nitration in the striatum (**c**) and cortex (**d**) in comparison with wild-type mice. Note the increased basal level of protein oxidation in the *Cse*$^{-/-}$ mice.

LETTER

doi:10.1038/nature13134

Nuclear reprogramming by interphase cytoplasm of two-cell mouse embryos

Eunju Kang[1], Guangming Wu[2], Hong Ma[1], Ying Li[1], Rebecca Tippner-Hedges[1], Masahito Tachibana[1]†, Michelle Sparman[1], Don P. Wolf[1], Hans R. Schöler[2] & Shoukhrat Mitalipov[1]

Successful mammalian cloning using somatic cell nuclear transfer (SCNT) into unfertilized, metaphase II (MII)-arrested oocytes attests to the cytoplasmic presence of reprogramming factors capable of inducing totipotency in somatic cell nuclei[1–3]. However, these poorly defined maternal factors presumably decline sharply after fertilization, as the cytoplasm of pronuclear-stage zygotes is reportedly inactive[4,5]. Recent evidence suggests that zygotic cytoplasm, if maintained at metaphase, can also support derivation of embryonic stem (ES) cells after SCNT[6–8], albeit at low efficiency. This led to the conclusion that critical oocyte reprogramming factors present in the metaphase but not in the interphase cytoplasm are 'trapped' inside the nucleus during interphase and effectively removed during enucleation[9]. Here we investigated the presence of reprogramming activity in the cytoplasm of interphase two-cell mouse embryos (I2C). First, the presence of candidate reprogramming factors was documented in both intact and enucleated metaphase and interphase zygotes and two-cell embryos. Consequently, enucleation did not provide a likely explanation for the inability of interphase cytoplasm to induce reprogramming. Second, when we carefully synchronized the cell cycle stage between the transplanted nucleus (ES cell, fetal fibroblast or terminally differentiated cumulus cell) and the recipient I2C cytoplasm, the reconstructed SCNT embryos developed into blastocysts and ES cells capable of contributing to traditional germline and tetraploid chimaeras. Last, direct transfer of cloned embryos, reconstructed with ES cell nuclei, into recipients resulted in live offspring. Thus, the cytoplasm of I2C supports efficient reprogramming, with cell cycle synchronization between the donor nucleus and recipient cytoplasm as the most critical parameter determining success. The ability to use interphase cytoplasm in SCNT could aid efforts to generate autologous human ES cells for regenerative applications, as donated or discarded embryos are more accessible than unfertilized MII oocytes.

We studied messenger RNA expression levels and cellular localization of several maternal and embryonic factors in unfertilized oocytes and pre-implantation-stage embryos[9–13], namely, *Hsf1*, *Brg1*, *Bmi1* (also known as *Smarca4*), *Oct4* (also known as *Pou5f1*), *Sall4*, *Esrrb*, *Apobec1*, *Aid* (also known as *Aicda*) and *Tet1*. We initially normalized *Gapdh* expression and confirmed that mRNA levels were statistically similar in intact and enucleated embryos and that protein was evenly distributed in nuclei and cytoplasm[14] (Extended Data Fig. 1a–c). No significant differences were found in expression levels of these genes between intact and enucleated interphase zygotes and I2C (Extended Data Fig. 1d). Bmi1, Hsf1 and Brg1 proteins were also equally distributed throughout the cells and, therefore, enucleation does not seem to deplete these factors in the cytoplasm (Extended Data Fig. 2a, b).

Success in mammalian SCNT has been attributed to the use of G0/G1-arrested donor nuclei with mature, unfertilized oocytes naturally arrested at MII as the recipient cytoplasm[1,15]. The slight cell cycle mismatch in this case could presumably be corrected shortly after SCNT by nuclear envelope breakdown followed by premature chromosome condensation induced by metaphase-specific factors present in the cytoplast[16]. Thus, both the donor nucleus and recipient cytoplasm resume coordinated embryonic cell divisions after artificial activation of SCNT embryos. In clarifying the importance of cell cycle matching to reprogramming success, we established the timing of cleavage initiation, from which the cell cycle of the recipient I2C cytoplasm could be assessed. Then, we carefully timed the onset and progression of the mitotic cell cycle during the transition from zygote to two-cell-stage embryo. Most zygotes entered first mitosis between 29 and 35 h after human chorionic gonadotropin (hCG) administration and formed centrally localized metaphase spindles detectable with polarizing microscopy. Zygotes progressed quickly through anaphase and telophase, culminating in cell division and formation of the two-cell embryo. Approximately 30 min after the onset of cleavage, two-cell embryos formed nuclei that were visible microscopically, corroborated by nuclear envelope detection using lamin B immunocytochemistry. Nuclei became more prominent by 60 min after cleavage and increased in size during the next 10 h (Fig. 1a–c). Incorporation of 5-bromodeoxyuridine (BrdU), an indicator of S phase, was first detected approximately 3–4 h after cleavage onset and was evident in embryos up to 7–8 h after cleavage. Embryos labelled from 8 h after cleavage did not incorporate BrdU, suggesting their transition into G2 phase (Fig. 1b). The majority of mouse two-cell embryos completed the full cell cycle and entered into the next metaphase approximately 18–20 h after first cleavage. Thus, we determined the complete cell cycle of a mouse two-cell embryo, starting with the initial cleavage (0 h) and onset of G1 phase. G1 phase concluded and DNA synthesis and S phase were initiated approximately 3 h after the end of the previous metaphase. S phase ended by 8 h and blastomeres entered G2 phase, which lasted a minimum of 9 h (Fig. 1d). We next determined the cell cycle characteristics of two nuclear donor cell types—fetal fibroblasts (FFs) and ES cells—and sorted populations for G0/G1-, G2/M- and S-phase cells (Extended Data Fig. 3).

We then examined SCNT into enucleated I2C (Fig. 2a). When G0/G1-phase FFs were introduced into early G1 (0.5–1 h) enucleated blastomeres, 38% of reconstructed SCNT embryos progressed to blastocysts. This rate was comparable with conventional SCNT of G0/G1 FFs into enucleated MII oocytes (40%), although significantly reduced compared with fertilized controls (94%; Table 1). When G0/G1 FFs were fused into intermediate (1–2 h) or late (2–3 h) G1 enucleated blastomeres, blastocyst development significantly declined (Table 1) and transfer of G0/G1 FFs into S-phase I2C cytoplasm (4–6 h) or metaphase zygotes (0 h) resulted in developmental arrest. When actively proliferating FFs in S phase were introduced into S-phase blastomeres, 24% of the resultant SCNT embryos reached the blastocyst stage. We also used ES cells and cumulus cells for nuclear transfer and observed considerable blastocyst development when our matching criteria were satisfied (Table 1). To examine a possible role for sperm-activated cytoplasm in interphase SCNT success, we generated enucleated parthenogenetic I2C and carried out SCNT with FFs, observing blastocyst development comparable with SCNT into fertilized counterparts or into

[1]Division of Reproductive and Developmental Sciences, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, Oregon 97006, USA. [2]Max Planck Institute for Molecular Biomedicine, Münster 48149, Germany. †Present address: South Miyagi Medical Center, Miyagi 989-1253, Japan.

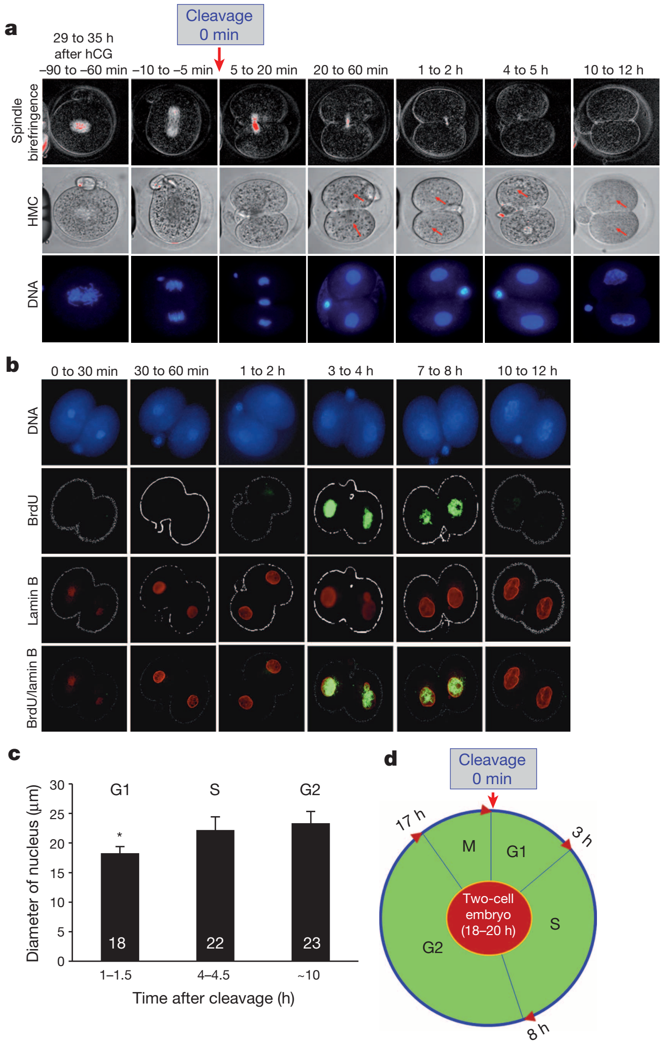


Figure 1 | Cell cycle progression in two-cell mouse embryos. **a**, Spindle and nuclear changes as a function of time after cleavage. The zygotic embryo began cleavage after going through a short telophase. We defined this event as 0 h in the cell cycle. The spindle disappeared gradually followed by nucleus formation and a visible increase in size. DNA images were produced from embryos that were fixed and stained with 4′,6-diamidino-2-phenylindole (DAPI), whereas spindle birefringence and Hoffman modulation contrast (HMC) images were taken from live embryos (original magnification, ×400). **b**, S-phase determination in two-cell embryos using BrdU incorporation and lamin B staining for identification of nuclear membrane (original magnification, ×400). **c**, Measurements of nuclear size as an indication of cell cycle progression ($n = 5$ biological replicates, $*P < 0.05$). Average diameter is shown ± standard deviation (s.d.). **d**, The full cell cycle timing in the two-cell mouse embryo.

conventional MII oocytes (Table 1). Thus, both artificial and sperm-activated cytoplasm of I2C support reprogramming. The arrest of cell-cycle-mismatched SCNT embryos could be secondary to nuclear arrest or apoptosis at the G1/S or S/G2 cell cycle checkpoints[17] (Extended Data Fig. 4).

We used FFs carrying the green fluorescent protein (GFP) transgene under the control of the *Oct4* promoter[18] and β-galactosidase[19] to confirm the origin of the nucleus in cloned embryos. The GFP signal was not detectable after fusion with I2C cytoplasm (Fig. 2b). Re-expression of the GFP transgene was observed in late four-cell embryos that continued into the compact morula. As expected, GFP expression was confined to the inner cell mass (ICM), having disappeared in the trophectoderm (Fig. 2b). Nuclear donor FFs weakly expressed Brg1 (Extended Data Fig. 2b, c) and a similar signal was found in the nuclei of two-cell SCNT

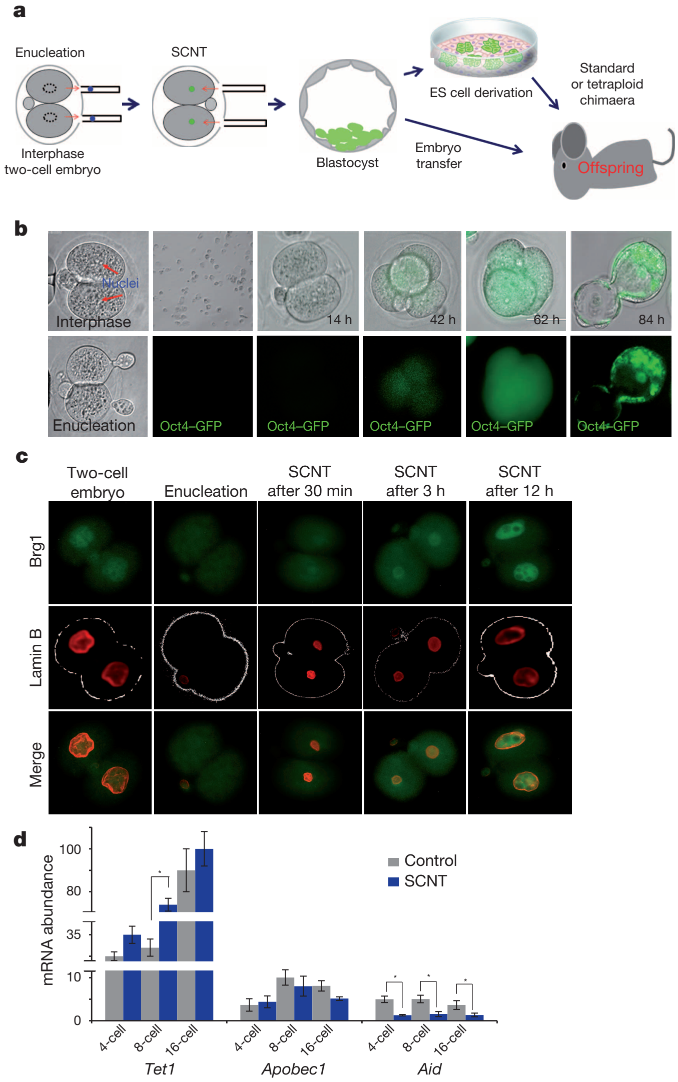


Figure 2 | Development of mouse embryos generated by interphase SCNT. **a**, Experimental design. Interphase SCNT conducted with two-cell embryos followed by ntES cell derivation or production of chimaeric or cloned offspring. **b**, Oct4–GFP re-expression after interphase SCNT using transgenic FFs (original magnification, ×400). **c**, Brg1 and lamin B protein expression in SCNT embryos (original magnification, ×400). **d**, Expression of *Tet1*, *Apobec1* and *Aid* in control and SCNT embryos ($n = 3$ technical replicates, $*P < 0.05$). Data show average ± s.d.

embryos shortly after introduction of FF nuclei into I2C cytoplasm (Fig. 2c). However, Brg1 protein became increasingly detectable in the nuclei of SCNT embryos, 3 and 12 h after transfer, and in the four- and eight-cell stages (Fig. 2c and Extended Data Fig. 2d). Whereas Hsf1 was not detected in FFs, it was expressed in SCNT embryos (Extended Data Fig. 2d). To examine expression of candidate reprogramming factors in SCNT embryos generated with I2C cytoplasm, we also measured mRNA levels for *Apobec1*, *Aid* and *Tet1*. Expression patterns and levels of *Apobec1* and *Tet1* were comparable with fertilized controls ($P < 0.05$), with the exception of the eight-cell stage, at which *Tet1* was upregulated in SCNT embryos. *Aid* transcript levels were significantly lower in SCNT embryos compared with controls (Fig. 2d).

The developmental potential of embryos produced by SCNT with FFs into enucleated I2C was evaluated further by plating blastocysts ($n = 26$) onto feeder layers. All adhered, eight formed typical ICM outgrowths and, after subsequent passaging, seven stable nuclear transfer (nt)ES cell lines (27%) with typical colony morphology and expression of Oct4 and SSEA-1 (also known as Fut4) were established (Fig. 3a and

Table 1 | *In vitro* development of embryos after nuclear transfer into enucleated I2C

Recipient cytoplasm type	Recipient cytoplasm cell cycle (hours after cleavage)	Nuclear donor cell cycle	Donor cell type	N	Cleaved embryos (%)	Blastocysts (%)
In vivo fertilized two-cell embryos	Metaphase	G0/G1	FFs	55	2 (4)	0
	Early G1 (0.5–1)	G0/G1	FFs	227	110 (48)	86 (38)
	Intermediate G1 (1–2)	G0/G1	FFs	56	12 (21)	8 (14)
	Late (2–3)	G0/G1	FFs	37	5 (14)	4 (11)
	S (4–6)	S	FFs	72	27 (38)	17 (24)
	S (4–6)	G0/G1	FFs	34	0	0
	Early G1 (0.5–1)	G1	ES cells	161	63 (39)	53 (33)
	Early G1 (0.5–1)	S	ES cells	43	0	0
	S (4–6)	S	ES cells	744	335 (45)	293 (39)
	S (4–6)	G1	ES cells	34	3 (9)	2 (6)
	Early G1 (0.5–1)	G0/G1	CCs	257	58 (23)	39 (15)
	Intact controls	N/A	N/A	135	133 (99)	127 (94)*
Parthenogenetically activated two-cell embryos	Early G1 (0.5–1)	G0/G1	FFs	32	15 (47)	11 (34)
	Intact parthenote controls	N/A	N/A	86	84 (98)	78 (91)*
MII oocytes	MII	G0/G1	FFs	289	161 (56)	116 (40)
	MII	G1	ES cells	113	46 (41)	25 (22)
	MII	G0/G1	CCs	136	95 (70)	23 (17)
	MII	S	ES cells	70	8 (11)	2 (3)

Control versus experimental treatment within each group. CCs, cumulus cells; N/A, not applicable.
* $P < 0.05$.

Extended Data Table 1). Similarly, three ntES cell lines were generated from 14 blastocysts (21%) after interphase SCNT with cumulus cells (Extended Data Table 1). With FFs carrying *Oct4-GFP* as nuclear donor cells, the resulting ntES cells re-expressed GFP at high levels (Fig. 3a). We confirmed normal karyotype and gender in all ntES cell lines (Extended Data Table 1) and corroborated their nuclear DNA origin from somatic cells and mitochondrial (mt)DNA inheritance from I2C cytoplasm (Fig. 3b, Extended Data Fig. 5a, b and Extended Data Table 1). The ability to form chimaeric offspring was evaluated by random selection of female ntES cells derived from FFs (ntES2) and injection into diploid eight-cell embryos. After transfer into recipients, 14 pups were born, 10 of which were chimaeric on the basis of the coat colour contribution of ntES cells on an albino background of host ICR embryos (Extended Data Table 2). The chimaera rate varied from 5% to more than 90% as determined by visual evaluation. Pups produced by breeding chimaeric females with control males showed germline transmission of the ntES cell genotype (Extended Data Fig. 5c).

We tested ntES cells for the ability to generate all-ES-cell offspring by tetraploid complementation[20]. After transfer of 165 tetraploid embryos injected with ntES2 cells, five live pups were recovered, of which four survived to adulthood and exhibited exclusively the ntES-cell coat colour phenotype (Extended Data Fig. 5d and Extended Data Table 2). We corroborated the all-ntES-cell origin of these mice by microsatellite and mtDNA genotyping (Extended Data Fig. 5a, b, e). Quantitative mtDNA analysis was characterized by the absence of any detectable contribution from host tetraploid embryos (Extended Data Fig. 5b). Additionally, two ntES cell lines derived from adult cumulus cells contributed to chimaeras using both diploid and tetraploid complementation assays (Fig. 3c, d and Extended Data Tables 1, 2). These stringent pluripotency tests confirm that ntES cells derived by SCNT into enucleated I2C blastomeres are the equivalent of embryo-derived ES cells.

Embryos produced by nuclear transfer into I2C cytoplasm were evaluated for totipotency by direct transfer into uteri of pseudopregnant females. When 115 SCNT blastocysts generated with FFs were transferred into ten recipients, four became pregnant and showed implantation sites, but none contained viable fetuses (Extended Data Table 3). An explanation for failed pregnancy may be the absence of a histone deacetylase inhibitor, which is essential for efficient live offspring production in conventional SCNT with MII oocytes[21]. However, when 89 nuclear transfer blastocysts generated with ES cell nuclei carrying a *lacZ* transgene were transferred, four live pups were recovered by caesarean section (Extended Data Fig. 6a and Extended Data Table 3). Their body and placental weights varied significantly (Extended Data Fig. 6b). Whole-body 5-bromo-4-chloro-3-indoyl-β-D-galactoside (X-gal) staining confirmed the nuclear transfer origin of the offspring (Fig. 3f) and placenta

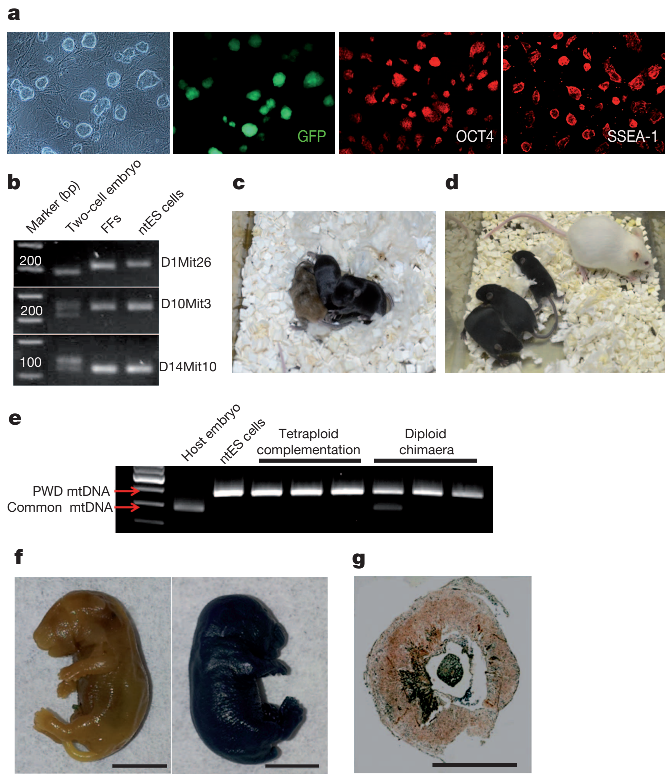


Figure 3 | Characterization of ntES cells and live offspring born after interphase SCNT. **a**, Morphology of ntES cells and expression of the pluripotency markers Oct4 and SSEA-1 by immunostaining. Second from the left, re-expression of GFP in ntES cells derived from FFs carrying Oct4–GFP (original magnification, ×100). **b**, Genotyping of ntES cells using chromosomal markers D1Mit26, D10Mit3 and D14Mit10. Numbers indicate the position of the chromosome. **c**, Diploid chimaeric pups generated by injection of ntES cells from cumulus cells (black) into ICR host embryos (albino). Note that two of the three pups contained exclusively the ntES cell phenotype. **d**, Tetraploid complementation pups produced after injection of ntES cells (black) into tetraploid host embryo (agouti). **e**, Genotyping of chimaeric offspring generated with ntES cells derived from cumulus cells by mtDNA profiling. All tetraploid and two of the three diploid chimaeras contained ntES cell mtDNA only. PWD is mtDNA haplotype derived from PWD/ph mice (GenBank accession EF108343.1). **f**, X-gal staining of control (non-stained, left) and cloned pup carrying the *lacZ* gene (blue stained, right). Scale bar, 0.5 cm. **g**, X-gal detection in placental tissue section recovered from a cloned pup. Scale bar, 0.5 cm.

(Fig. 3g). In addition, cloned offspring and their placentas carried *lacZ* and *GFP* transgenes (Extended Data Fig. 6c).

Interphase nuclear transfer into fertilized embryos has resulted in live offspring when pronuclei were exchanged between two zygotes[4] and when four- or eight-cell blastomere nuclei were transferred into enucleated two-cell embryos[22]. However, nuclear transfer of more advanced ICM cells, ES cells or somatic cells into interphase cytoplasm resulted in early developmental arrest, probably reflecting mismatched cell cycles[5,7,22,23]. Nuclear transfer of 4–16-cell blastomeres into enucleated MII oocytes has also resulted in live offspring in several species, including nonhuman primates[24–27]. This approach required pre-activation of the recipient cytoplast before nuclear transfer, resulting in exit from the metaphase and progression to the interphase[24,28]. Because most randomly chosen nuclear donor blastomeres were also in interphase, synchronization between the nucleus and cytoplast was probably achieved.

The present findings suggest that the interphase cytoplasm has reprogramming capacity and that precise cell cycle synchrony between donor and recipient can lead to the development of functional ES cells or live offspring. This supposition would undermine the premise that SCNT failures involving interphase cytoplasm are secondary to the absence of critical maternal factors removed during enucleation. In view of the present results in mouse and the unsuccessful attempt to conduct SCNT into human metaphase zygotes[29], SCNT into human two-cell embryos with proper cell cycle matching should be reconsidered.

METHODS SUMMARY

All animal procedures were approved by the Institutional Animal Care and Use Committee at Oregon Health & Science University. Two-to-three-month-old B6D2F1 (C57BL/6J female × DBA/2 male) or B6D2F1-mt$^{PWD/ph}$ (C57BL/6J-mt$^{PWD/ph}$/ForeJ female × DBA/2 male) females were superovulated with 5 international units (IU) pregnant mare's serum gonadotropin (PMSG) and hCG and mated to B6D2F1-mt$^{PWD/ph}$ males. Zygotes were collected from the excised oviducts of plugged females at 24–26 h after hCG injection and monitored for cleavage. Nuclear transfer was carried out based on timing after the first cleavage division, 30–38 h after hCG treatment.

Nuclear donor FFs, cumulus cells or ES cells suspended in Dulbecco's phosphate buffered saline (DPBS) (10 μl) were mixed with 5 μl haemagglutinating virus of Japan envelope (HVJ-E; GenomONE-CF HVJ Envelope Cell Fusion Kit) and kept on ice until use. Modified I2C enucleation and nuclear transfer were done as shown in Supplementary Video 1. SCNT embryos were placed into droplets of KSOM medium covered with mineral oil and cultured to the blastocyst stage. Conventional MII SCNT was performed as described[21] except that enucleation was done with spindle observation using an optical birefringence system (OosightTM; http://www.cri-inc.com) in a glass-bottom dish. All manipulations were completed on the heated stage (37 °C) of an Olympus microscope equipped with a XYClone laser (http://www.hamiltonthorne.com) for zona drilling.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 November 2013; accepted 6 February 2014.
Published online 26 March 2014.

1. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385,** 810–813 (1997).
2. Gurdon, J. B. & Wilmut, I. Nuclear transfer to eggs and oocytes. *Cold Spring Harb. Perspect. Biol.* **3,** a002659 (2011).
3. Tachibana, M. *et al.* Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153,** 1228–1238 (2013).
4. McGrath, J. & Solter, D. Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development *in vitro*. *Science* **226,** 1317–1319 (1984).
5. Wakayama, T., Tateno, H., Mombaerts, P. & Yanagimachi, R. Nuclear transfer into mouse zygotes. *Nature Genet.* **24,** 108–109 (2000).
6. Egli, D., Sandler, V. M., Shinohara, M. L., Cantor, H. & Eggan, K. Reprogramming after chromosome transfer into mouse blastomeres. *Curr. Biol.* **19,** 1403–1409 (2009).
7. Egli, D., Rosains, J., Birkhoff, G. & Eggan, K. Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447,** 679–685 (2007).
8. Riaz, A. *et al.* Mouse cloning and somatic cell reprogramming using electrofused blastomeres. *Cell Res.* **21,** 770–778 (2011).
9. Egli, D., Birkhoff, G. & Eggan, K. Mediators of reprogramming: transcription factors and transitions through mitosis. *Nature Rev. Mol. Cell Biol.* **9,** 505–516 (2008).
10. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95,** 379–391 (1998).
11. Zhang, J. *et al.* Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nature Cell Biol.* **8,** 1114–1123 (2006).
12. Festuccia, N. *et al.* Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* **11,** 477–490 (2012).
13. Jullien, J., Pasque, V., Halley-Stott, R. P., Miyamoto, K. & Gurdon, J. B. Mechanisms of nuclear reprogramming by eggs and oocytes: a deterministic process? *Nature Rev. Mol. Cell Biol.* **12,** 453–459 (2011).
14. Zheng, L., Roeder, R. G. & Luo, Y. S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell* **114,** 255–266 (2003).
15. Wakayama, T., Rodriguez, I., Perry, A. C., Yanagimachi, R. & Mombaerts, P. Mice cloned from embryonic stem cells. *Proc. Natl Acad. Sci. USA* **96,** 14984–14989 (1999).
16. Mitalipov, S. M. *et al.* Reprogramming following somatic cell nuclear transfer in primates is dependent upon nuclear remodeling. *Hum. Reprod.* **22,** 2232–2242 (2007).
17. Jaroudi, S. & SenGupta, S. DNA repair in mammalian embryos. *Mutat. Res.* **635,** 53–77 (2007).
18. Schöler, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J.* **9,** 2185–2195 (1990).
19. Zambrowicz, B. P. *et al.* Disruption of overlapping transcripts in the ROSA βgeo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl Acad. Sci. USA* **94,** 3789–3794 (1997).
20. Wu, G. *et al.* Generation of healthy mice from gene-corrected disease-specific induced pluripotent stem cells. *PLoS Biol.* **9,** e1001099 (2011).
21. Kishigami, S. *et al.* Significant improvement of mouse cloning technique by treatment with trichostatin A after somatic nuclear transfer. *Biochem. Biophys. Res. Commun.* **340,** 183–189 (2006).
22. Tsunoda, Y. *et al.* Full-term development of mouse blastomere nuclei transplanted into enucleated two-cell embryos. *J. Exp. Zool.* **242,** 147–151 (1987).
23. Eckardt, S., Leu, N. A., Kurosaka, S. & McLaughlin, K. J. Differential reprogramming of somatic cell nuclei after transfer into mouse cleavage stage blastomeres. *Reproduction* **129,** 547–556 (2005).
24. Mitalipov, S. M., Yeoman, R. R., Nusser, K. D. & Wolf, D. P. Rhesus monkey embryos produced by nuclear transfer from embryonic blastomeres or somatic cells. *Biol. Reprod.* **66,** 1367–1373 (2002).
25. Stice, S. L. & Robl, J. M. Nuclear reprogramming in nuclear transplant rabbit embryos. *Biol. Reprod.* **39,** 657–664 (1988).
26. Willadsen, S. M. Nuclear transplantation in sheep embryos. *Nature* **320,** 63–65 (1986).
27. Cheong, H. T., Takahashi, Y. & Kanagawa, H. Birth of mice after transplantation of early cell-cycle-stage embryonic nuclei into enucleated oocytes. *Biol. Reprod.* **48,** 958–963 (1993).
28. Stice, S. L., Keefer, C. L. & Matthews, L. Bovine nuclear transfer embryos: oocyte activation prior to blastomere fusion. *Mol. Reprod. Dev.* **38,** 61–68 (1994).
29. Egli, D. *et al.* Reprogramming within hours following nuclear transfer into mouse but not human zygotes. *Nature Commun.* **2,** 488 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors would like to acknowledge the Small Lab Animal Unit at Oregon National Primate Research Center for providing expertise and services that contributed to this project. We are indebted to E. Wolff, A. Sugawara, C. Ramsey, H.-S. Lee, J. Woodward, D. Melquizo Sanchis, C. Van Dyken and B. Manson for their technical support. The study was supported by grants from the National Institutes of Health (R01HD063276, R01HD057121, R01HD059946, R01EY021214, P51OD011092) and funds from the Leducq Fondation and the Collins Medical Trust.

Author Contributions S.M. and E.K. conceived and planned the study, E.K. carried out SCNT experiments, G.W. conducted tetraploid chimera experiments, H.M., Y.L., R.T.-H., M.T. and M.S. assisted with colony management, oocyte and embryo collections, embryo transfers, ES cell isolation and culture and expression profiling. E.K., D.P.W., H.R.S. and S.M. analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M. (mitalipo@ohsu.edu).

METHODS

Preparation of nuclear donor cells. The carcass of an embryonic day 14 fetus was collected and treated with 0.1% collagenase for 30 min. The digested tissue was cultured with 10% FBS and 100 μg ml^{-1} penicillin/streptomycin in DMEM. Early passage (2–4) FFs were used for SCNT. For G0/G1 phase, cells were seeded in 4-well plates, and cultured for 1 week with 0.5% FBS medium. For S phase, G0/G1-arrested FFs were seeded at low density in ESC growth medium (see later), and cultured for 20 h. The cumulus–oocyte complexes were collected from superovulated females 18–20 h after hCG injection (see later) and briefly exposed to a medium containing 0.1% hyaluronidase for disaggregation. Dispersed cumulus cells were rinsed and kept in KSOM medium (Millipore) before use.

ES cells and FFs were fixed overnight in 70% ethanol, washed in DPBS and stained with 10 μg ml^{-1} propidium iodide (Sigma) in DPBS for 1 h on ice for flow cytometric analysis (BD Biosciences, LSRII flow cytometer). FFs were either arrested at G0/G1 by culture to confluency under low serum conditions or forced to proliferate and enter S phase after re-plating G0/G1-arrested cells at low density in medium containing serum. In confluent FFs, nearly 79% of the cells were at G0/G1, 3% were at S and 18% were at G2/M phase. However, after sorting by size (small ≤ 12 μm, large ≥16 μm), small-sized FFs were enriched for G0/G1 cells (91%). In proliferating FFs, the proportion of S-phase cells was significantly increased (70%). We also confirmed S phase by incorporation of BrdU (BD PharmingenTM; 30 min incubation at 30 μM) after 3.7% formaldehyde fixation. There was no significant correlation between cell size and cell cycle phase distribution in proliferating FFs. In actively proliferating ES cells, 27% were at G0/G1, 47% at S and 26% at the G2/M phase. When ES cells were sorted on the basis of size, approximately 70% of small cells (≤9 μm) were G0/G1. In contrast, most medium-sized cells (10–15 μm) were in S phase (BrdU^{+}) whereas a significant percentage of large cells (≥16 μm) were in G2/M phase. On the basis of these observations, most small-sized ES cells and confluent FFs were in G0/G1 phase, whereas proliferating FFs or medium-sized ES cells were predominantly in S phase.

Cell cycle determination in recipient embryos. Pronuclear-stage zygotes were collected from superovulated, mated females and continuously examined before and after the first cleavage division by HMC and polarized microscopy. Selected embryos were also fixed and labelled with antibodies against lamin B and DAPI to follow nuclear membrane and chromatin dynamics, respectively. To define the onset of S phase, we sampled every hour after the onset of cleavage into the two-cell stage with BrdU, which incorporates into newly synthesized DNA during replication. The embryos were incubated in 100 μM BrdU for 20 min with the remaining steps carried out according to the manufacturer's instructions before embryo observation by fluorescence microscopy (×400).

Interphase two-cell SCNT. Nuclear transfer of FFs, ES cells or cumulus cells into enucleated I2C was undertaken, ensuring cell cycle match between the recipient cytoplasm and the transplanted nucleus. *In vivo* fertilized, pronuclear-stage zygotes were recovered and continuously monitored for the onset of cleavage. Two-cell blastomeres enucleated within 0.5–3 h after the onset of cleavage were considered to be G1 cytoplasts whereas enucleations conducted 4–6 h after the first mitosis were defined as S phase. Both blastomeres were enucleated under HMC, and an intact nuclear donor cell was introduced into each enucleated blastomere using the fusogenic activity of HVJ-E (Fig. 2b and Supplementary Video 1).

Derivation and culture of ES cells. The zona pellucida was removed with acidic Tyrode's solution (Sigma-Aldrich) and denuded blastocysts were placed onto mitomycin C (Sigma) treated mouse embryonic fibroblast (mEF) feeder layers in ESC derivation medium: KO DMEM (Invitrogen) containing 15% KOSR (Invitrogen), 5% FBS (HyClone), mitogen-activated protein kinase inhibitor (PD98059; Cell Signalling) and 1,000 units ml^{-1} LIF (Sigma). After ICM outgrowth, the cultures were dispersed into single cells using trypsin and seeded on fresh mEFs in complete ESC medium: KO DMEM, 1 mM L-glutamine, 100 units ml^{-1} penicillin/100 μg ml^{-1} streptomycin, 100 μM β-mercaptoethanol (Sigma), 100 μM nonessential amino acids (Invitrogen), 1,000 units ml^{-1} LIF, 10% FBS and 10% KOSR.

Chimaera assay and embryo transfer. Approximately 10 ES cells were injected into 4–8-cell embryos recovered from B6D2F1 females mated with C3H males for tetraploid embryo complementation or from ICR for diploid chimaeras. Injected embryos were cultured for 1 day in KSOM. SCNT or chimaeric blastocysts were transferred into the uteri of pseudopregnant (E2.5) ICR females. Caesarean section was performed from embryonic day 18.5 to 20.5.

Genotyping analyses. Genotyping primers, D1Mit26, D10Mit3 and D14Mit10 were obtained from Mouse Genome Informatics (http://www.informatics.jax.org). *GFP*, *lacZ* and mitochondrial analysis primers are provided later. DNA was extracted from tail tips using Arcturus Microgenomics reagent kit (Applied Biosystems) or from cultured cells using a DNA kit (Gentra SYSTEMS). PCR products were separated by 3% agarose gels and visualized by ethidium bromide staining.

PCR products were sequenced and the informative single nucleotide polymorphic (SNP) sites were identified using Sequencher v.4.7 software (GeneCodes). Allele refractory mutation system quantitative PCR (ARMS-qPCR) assays were performed to measure mtDNA heteroplasmy levels in chimaeric offspring as previously described[30]. For the discriminative assay, primers and TaqMan probes (FAM) were designed to target to the unique mtDNA region in ND5 gene for the wild-type and PWD/ph mtDNA. For the consensus assay, primers and probes (TET) were targeted to the common region of the COXIII gene. Multiplex PCR reactions (15 μl) containing 1× PCR Master Mix, 100–250 nM of each set of primers, 150 nM of each TaqMan probe and 4 ng of total genomic DNA at a 1:8 dilution were run using an ABI 7900HT (Applied Biosystems).

X-gal staining. Fetuses and placenta (embryonic day 18.5) were removed by caesarean section and fixed overnight in 3.7% formaldehyde at 4 °C, washed three times in PBS, cleared in 70% ethanol followed by incubation in 30% sucrose at 4 °C overnight, embedding (Tissue-Tek OCT compound; Sakura Finetechnical) and freezing on dry ice. Sagittal sections (10 μm) were cut at −18 °C with a cryostat (Leica Instrument). The specimen was stained overnight at 37 °C in X-gal (1 mg ml^{-1}), potassium ferricyanide (5 mM), potassium ferrocyanide (5 mM), $MgCl_2$ (2 mM) in PBS.

Analysis of gene expression. Total RNA was isolated and pooled from ten embryos for each experiment (MII oocytes (18 h), interphase (one cell, 24–26 h), metaphase (one cell, 28–30 h), two cell (31–33 h), two-cell cytoplast (nuclei removed, 31–33 h), four cell (50–52 h), eight cell (66–68 h), 16 cell (74–76 h) and blastocyst (98–100 h)) with an Arcturus pico pure RNA isolation kit (Life Technologies). RNA was immediately treated with DNase I (Invitrogen) and reversed into cDNA using the SuperScript III first-strand synthesis system (Invitrogen) according to the manufacturer's instructions. RT–PCR was performed using an ABI 7900HT Real time PCR system and the SYBR Green PCR mix (Applied Biosystems). The samples were heated to 95 °C for 10 min and run for 40 cycles of: 94 °C for 15 s, 58 °C for 20 s, 72 °C for 30 s. Gene expression was normalized to *Gapdh*.

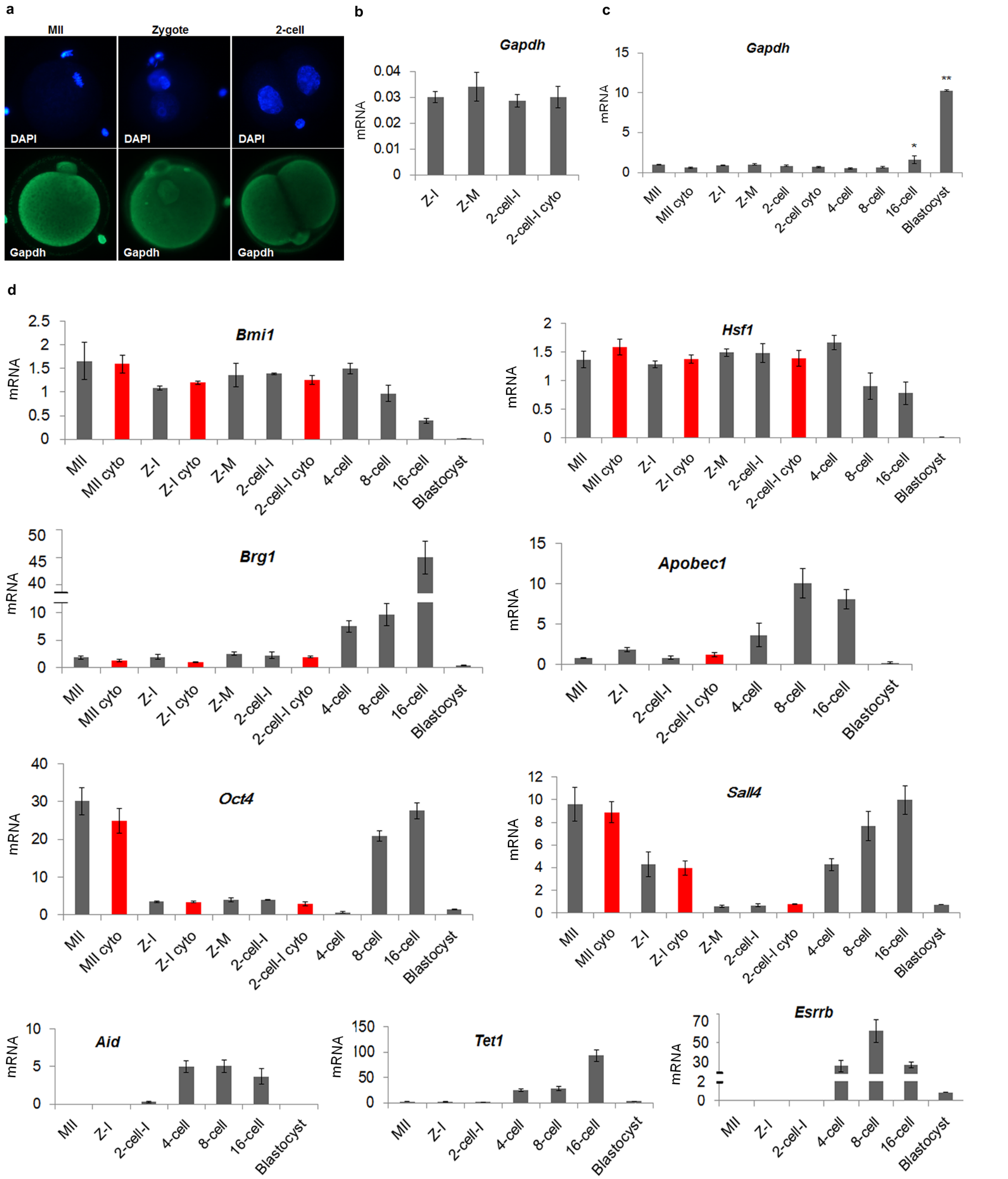
Immunostaining. Embryos were labelled with antibodies for Brg1 (SC-10768), lamin B (SC-6217), Bmi1 (SC-30943) and Hsf1 (SC-9144). ES cells were stained with antibodies for Oct4 (SC-30943) and SSEA-1(SC-21702). Secondary antibodies were coupled with FITC or Cy3 and DAPI was used for counterstaining nuclei. All antibodies were purchased from Santa Cruz.

Western blot. Forty embryos were collected and washed three times in TBS and transferred into 7 μl of RIPA extraction buffer supplemented with protease inhibitors. Protein samples were diluted with 2.5 μl sample buffer and 1 μl reducing agent, incubated for 10 min at 70 °C, and size separated using 4–12% Bis-Tris Mini Gel (novex; Life Technologies). Proteins were blotted into polyvinylidene fluoride (PVDF) membranes (Hybond-P; GE Healthcare). Nonspecific binding was blocked by blocking solution (Invitrogen) at room temperature for 1 h. Membranes were probed for 1 h at room temperature with primary antibody (1:200 rabbit anti-Brg1, SC-10768; 1:5,000 rabbit anti-Gapdh, G9545; Sigma). Primary antibodies were detected with an HRP-conjugated secondary antibody (1:10,000 goat anti-rabbit) applied for 1 h at room temperature and developed with Femto Maximum Chemiluminescent Substrates (SuperSignal 34095; Thermo Scientific).

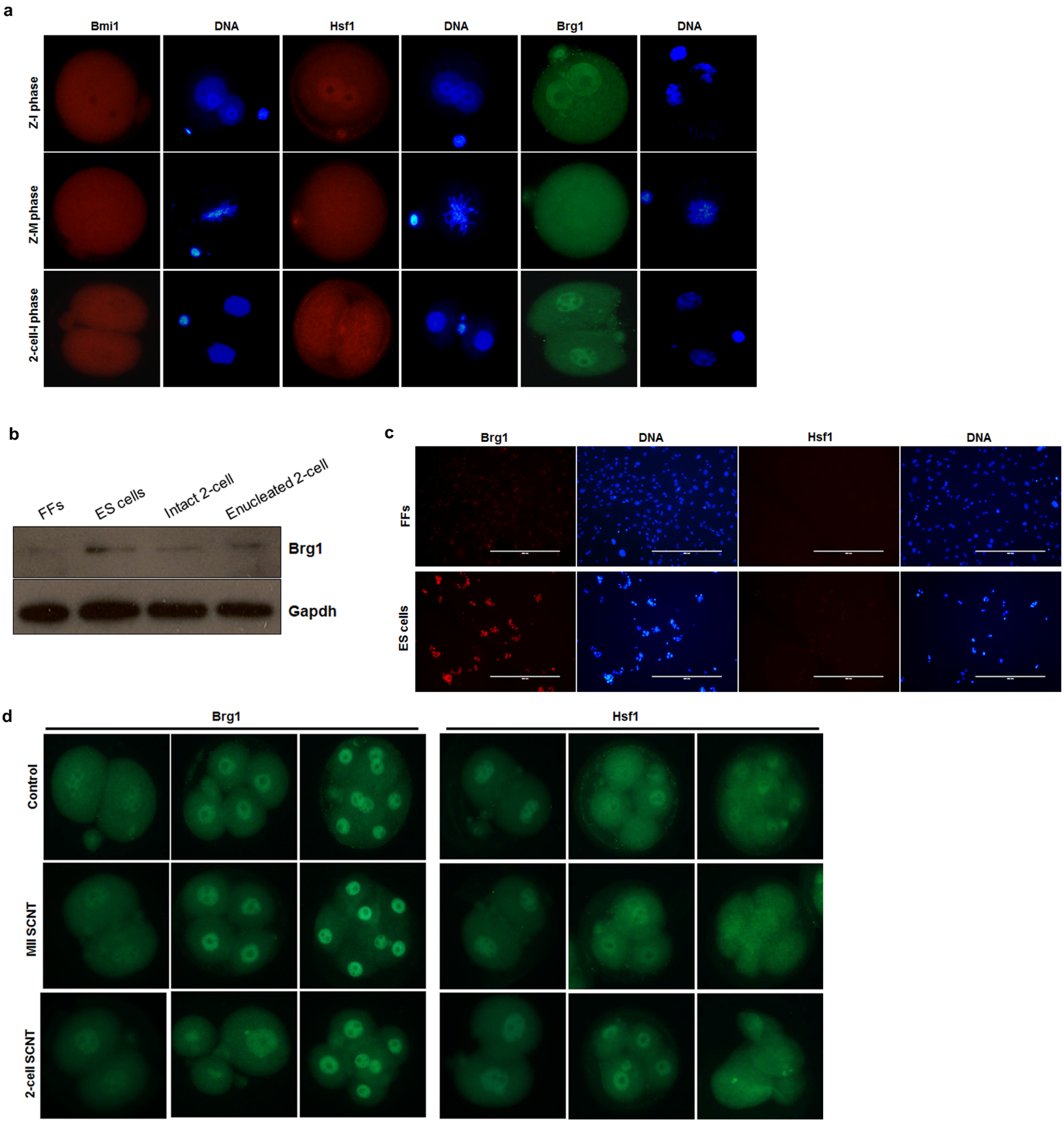
Statistical analysis. All data are presented as average ± s.d. One-way analysis of variance was used to compare differences among groups followed by post comparison between groups by Bonferroni using IBM SPSS statistics 20 software.

Primer information. All primers are given as 5′–3′. *Actb*, F-AGCCATGTACGTAGCCATCC, R-CTCTCAGCTGTGGTGGTGAA; *Gapdh*, F-CCACCCAGAAGACTGTGGAT, R-CACATTGGGGGTAGGAACAC; *Brg1*, F-CTGCGTAAGATCTGCAACCA, R-TTGGCAAAAGAGGAGCACTT; *Bmi1*, F-TGTCCAGGTTCACAAAACCA; R-TGCAACTTCTCCTCGGTCTT; *Hsf1*, F-CCAGCAGCAAAAAGTTGTCA, R-GTAGGCTGGAGATGGAGCTG; *Aid*, F-GGTACCTGGGAGTCGTTTGA, R-GGTGAATGCCACTTTCCCTA; *Apobec1*, F-AAGGATACCACCCCATCTCC, R-CCCTACCCATGAATGACACC; *Tet1*, F-GAGCCTGTTCCTCGATGTGG, R-CAAACCCACCTGAGGCTGTT; *Oct4*, F-GACTCCACCTCACACGGTTC, R-CAGACCACCATCTGTCGCT; *Sall4*, F-CGACCACCCAAGTATTGCCAG, R-AGGTGCAACAAGTCAGAGGAA; *Esrrb*, F-TGGACTCGCCGCCTATGTTCG, R-ACTTGCGCCTCCGTTTGGTGA; ROSA26-*lacZ*, F-CTTGTGATCCGCCTCGGAGTATT, R-CGCGCCGCTGTAAAGTGTTACGT; *GFP*, F-TGCAGTGCTTCAGCCGCTAC, R-TCGCCCTCGAACTTCACCTC; mtCO3, F-TCTAGCCTCGTACCAACACATGAT, R-TGAAACACCTGATGCTAGAAGTACTGA (ARMS qPCR); mtPWD/ph-ND2, F-ACTGCACATAGGACTTATTCTTGTC, R-TTGAGTAGCGGGTAGATTAGG (ARMS qPCR); mtWild-ND5, F-CCTACTAATTACACTAATCGCCACT, R-GAGGTCTGGGTCATTTTCGTTA (ARMS qPCR); mtCommon-ND5, F-TGCAACACCAACGCCTGAGCC, R-TTGTGTGAAGAGTTGAGGTGG; mtPWD/ph-ND5, F-CTAATCACACTAATTGCCACC, R-GTAGTAAGTGGCGTAATGTGG; mtWild-ND5, F-TTCCACCAACCAACATTCCAA, R-ATAATAAGTGGCGTAATGTGG.

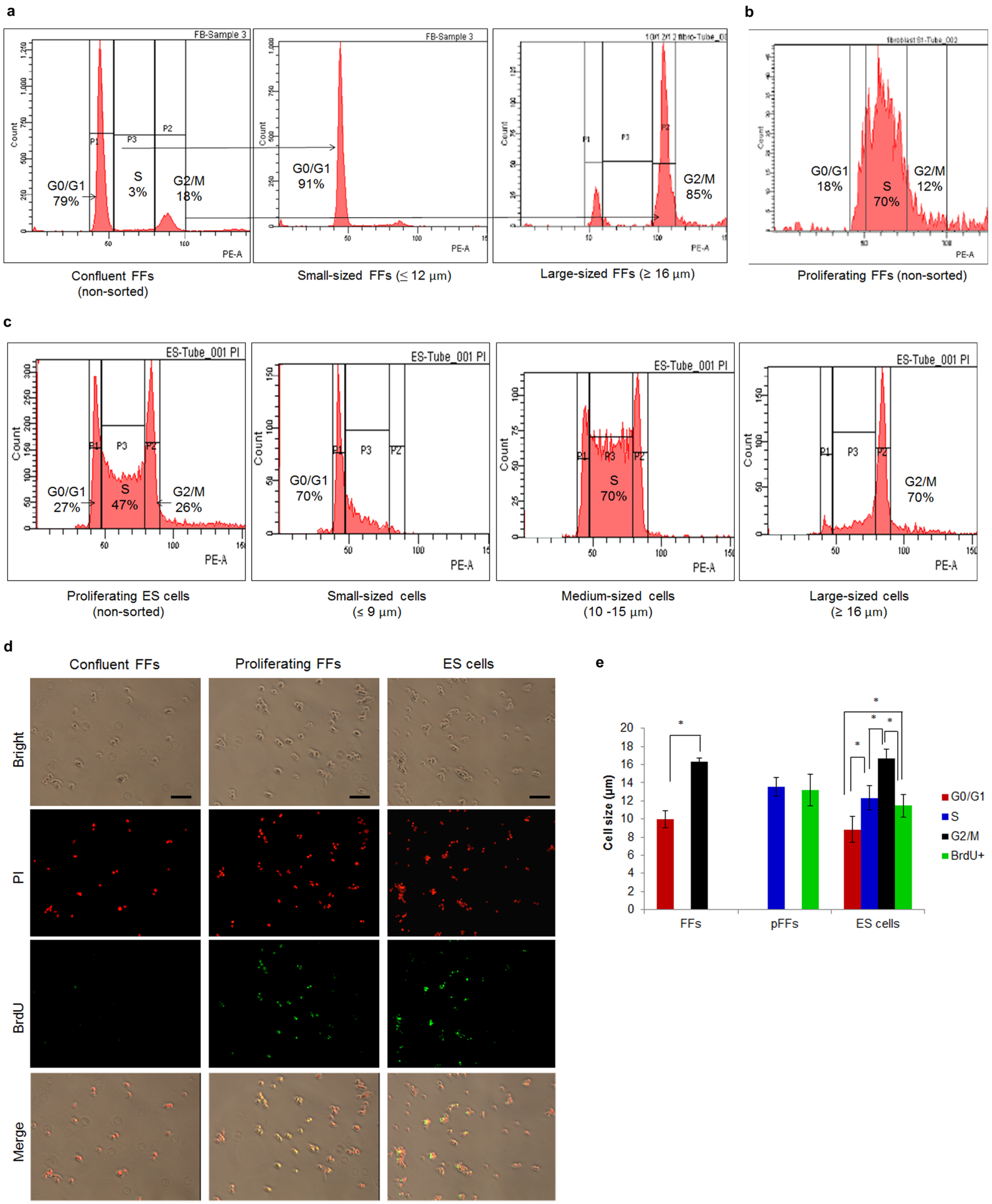
30. Lee, H. S. *et al.* Rapid mitochondrial DNA segregation in primate preimplantation embryos precedes somatic and germline bottleneck. *Cell. Rep.* **1,** 506–515 (2012).



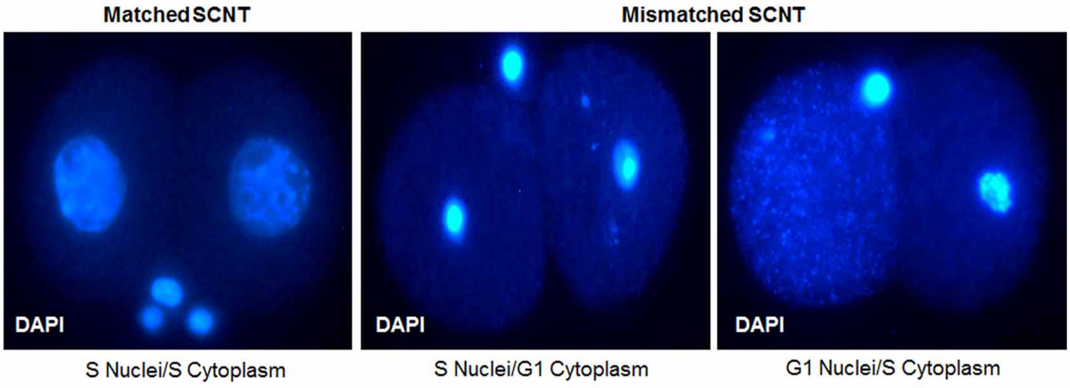
Extended Data Figure 1 | Maternal and embryonic gene expression patterns. **a**, Immunocytochemical detection of Gapdh signal demonstrated even distribution in nuclei and cytoplasm in MII oocytes and interphase zygotes and two-cell embryos (original magnification, ×400). **b**, Expression of *Gapdh* normalized to *Actb*. No significant differences were seen between intact and enucleated two-cell embryos (four replicates each containing pooled RNA from five embryos, $P > 0.05$). **c**, *Gapdh* expression was relatively low and constant with no significant differences seen until the eight-cell stage. Expression was increased in 16-cell embryos and blastocysts (three replicates each containing pooled RNA from ten embryos, $^{*}P < 0.05$; $^{**}P < 0.01$). **d**, Gene expression of the transcriptional regulators *Bmi1, Hsf1, Brg1, Sall4* and *Esrrb*, the transcription factor *Oct4*, and epigenetic factors *Apobec1, Aid* and *Tet1* during mouse pre-implantation embryo development. The level of expression of *Brg1*, *Apobec1*, *Oct4*, *Sall4*, *Aid*, *Tet1* and *Esrrb* underwent marked increases at or after the four-cell stage. The red bars indicate enucleated oocytes, zygotes or two-cell embryos. No significant differences were observed between intact and enucleated counterparts ($n = 3$ biological replicates, $P > 0.05$). Gene expression was normalized to *Gapdh*. Error bars indicate average ± s.d. cyto, cytoplast; I, interphase; M, metaphase; MII, metaphase II-arrested oocyte; Z, zygote.



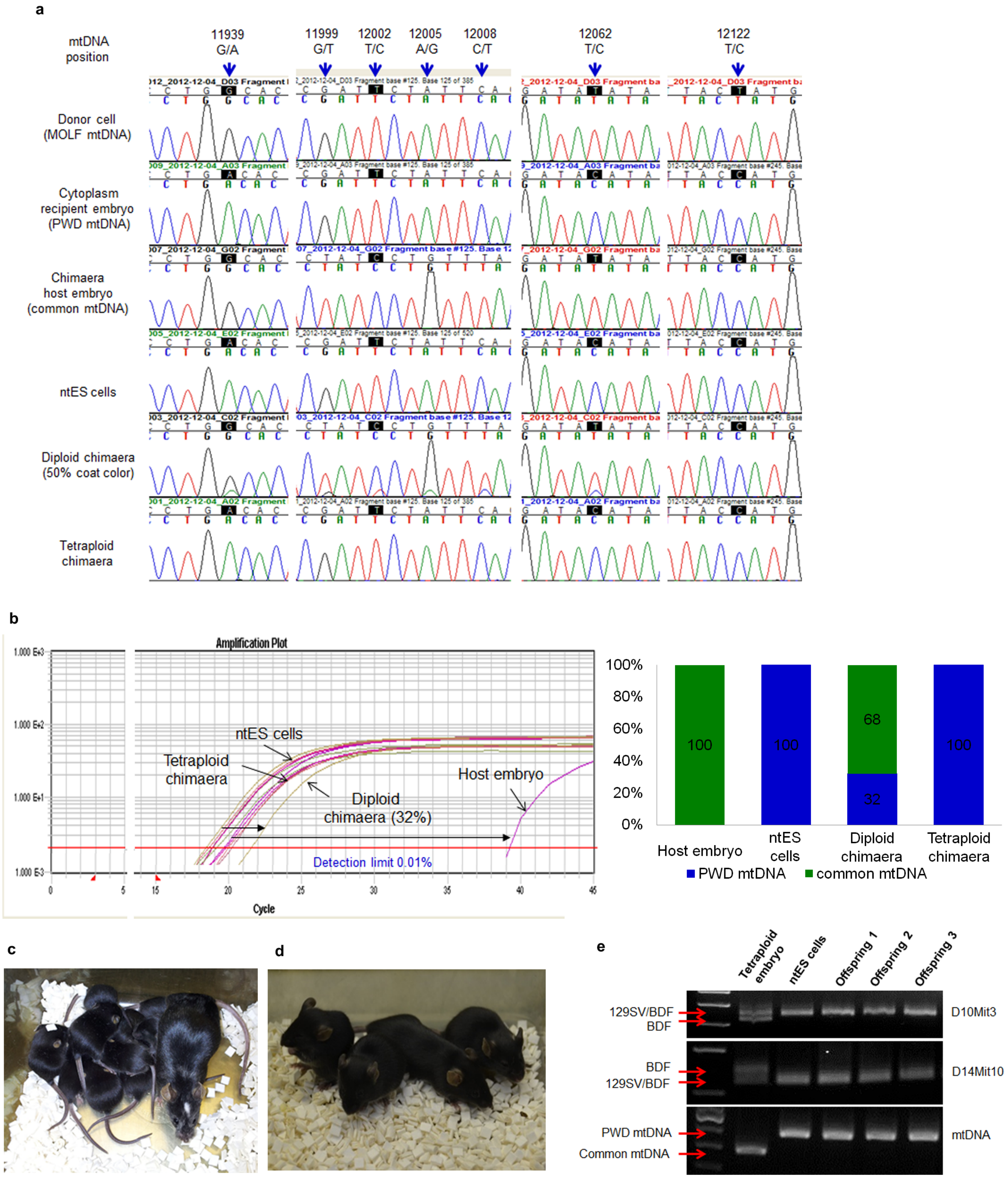
Extended Data Figure 2 | Protein expression of candidate reprogramming factors. **a**, Immunocytochemical detection of Bmi1, Hsf1 and Brg1 in zygotic interphase, metaphase and two-cell interphase embryos. The expression pattern of these proteins was similar to that observed for mRNA expression (Extended Data Fig. 1d). I2C embryos were fixed 4 h after cleavage. I, interphase; M, metaphase; Z, zygote. **b**, Western blot detection of Brg1 in nuclear donor cells and in intact and enucleated two-cell embryos ($n = 40$). **c**, Immunocytochemical detection of Brg1 and Hsf1 in nuclear donor FFs and ES cells. Scale bars, 400 μm. **d**, Immunocytochemical detection of Brg1 and Hsf1 in SCNT embryos (original magnification, ×400). MII, MII oocyte.



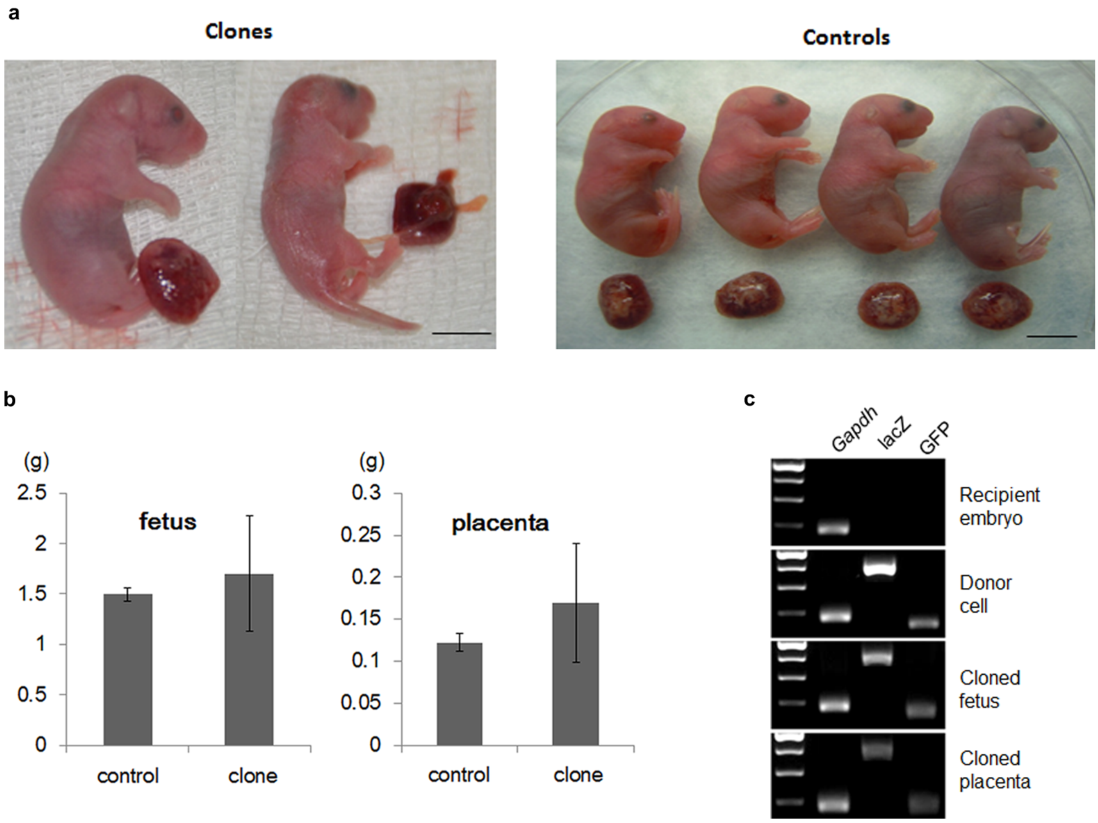
Extended Data Figure 3 | Cell cycle determination in nuclear donor cells. **a**, Confluent FF populations subjected to flow cytometric processing. After sorting by size and density, 91% of small-sized cells were in the G0/G1 phase of the cell cycle. **b**, Proliferating FF populations were mostly at S phase. **c**, ES cell populations by flow cytometry. The size was correlated with the cell cycle, 70% of the small-, medium- and large-sized cells were at G0/G1, S and G2/M, respectively. **d**, FITC-conjugated BrdU was used to define S-phase FFs and ES cells. PI, propidium iodide. Scale bars, 100 μm. **e**, Correlation of cell size and cell cycle in FFs and ES cells. Cell size was measured photographically after 3.7% formaldehyde fixation. S-phase cells were separated by BrdU integration (green). The assignments of cell size provided in panels **a**, **b** and **c** were defined from the results in panel **e** (*$P < 0.05$). pFFs, proliferating fetal fibroblasts.



Extended Data Figure 4 | Nuclear staining of cell-cycle-matched and mismatched SCNT embryos. Left, expanded nuclei of developmentally competent two-cell SCNT embryos generated by transfer of S-phase FF nuclei into enucleated S-phase two-cell embryo. Middle and right panels, condensed or dispersed nuclei of arrested SCNT embryos generated after cell cycle mismatch between donor nucleus and recipient cytoplasm. Embryos were arrested and apoptotic by the G1/S or S/G2 cell cycle checkpoints, respectively. SCNT embryos were fixed 15 h after cleavage (original magnification, ×400).



Extended Data Figure 5 | Genotyping of chimaeric mice. **a**, Chromatogram of mtDNA depicting sequence differences at 11,939, 12,062 and 12,122 base pairs (bp) and demonstrating that mtDNA in ntES cells was derived from recipient embryo cytoplasm. Sequence differences between chimaera host embryos and ntES cells at 11,939, 12,199, 12,002, 12,005, 12,008 and 12,062 bp allowed quantification of chimaerism. The mtDNA from a diploid chimaeric mouse appeared as a double peak. The tetraploid chimaera showed only ntES cell mtDNA (GeneBank accession EF108343.1). **b**, Allele refractory mutation system quantitative PCR (ARMS-qPCR) assay in diploid and tetraploid chimaeras. Similar colours indicate similar origin. Distances between colour plots represent the contribution of chimaeric host embryos. Amplification plot (left) and bar graph (right). Detection limit of this assay is 0.01%. **c**, High ntES cell contribution chimaeric female and her pups demonstrating germline transmission. White spot on the head of the dam originated from the ICR host embryo. **d**, Tetraploid embryo complementation pups born after ntES cell (black) injection into host eight-cell-stage tetraploid embryos. **e**, Genotyping of tetraploid complementation pups.



Extended Data Figure 6 | Cloned offspring and genotyping. **a**, Body and placenta of cloned and control pups delivered by caesarean section at 20 days. Clones and placentas were larger or smaller than controls. Scale bars, 0.5 cm. **b**, Fetal and placental weights (g) of cloned and control pups. **c**, Genotyping of nuclear donor cells and cloned pups carrying ROSA26$^{+/-}$OG2$^{+/-}$.

Extended Data Table 1 | Origin of ntES cell lines generated by SCNT into enucleated I2C

ntES cells	Gender	Nuclear donor cells	Genotype of donor cells	mtDNA of donor cells	2-cell embryo mtDNA	Diploid chimaera	Tetraploid chimaera
ntES1	M	FFs	B6D2F1	common	PWD	n/t	n/t
ntES2	F	FFs	129S1/SvImJ	MOLF	PWD	yes	yes
ntES3	M	FFs	B6D2F1	common	PWD	yes	n/t
ntES4	M	FFs	B6D2F1	common	PWD	n/t	n/t
ntES5	M	FFs	B6D2F1	common	common	n/t	n/t
ntES6	F	FFs	Oct4-GFP	common	common	n/t	n/t
ntES7	F	FFs	Oct4-GFP	common	PWD	n/t	n/t
ntES8	F	CCs	B6D2F1	PWD	PWD	yes	yes
ntES9	F	CCs	B6D2F1	PWD	PWD	n/t	yes
ntES10	F	CCs	B6D2F1	PWD	PWD	n/t	n/t

CCs, cumulus cells; n/t, not tested.

Extended Data Table 2 | Ability of ntES cells to contribute to chimaeras

Cell line	Host embryo type	N	N recipients (pregnant)	N Pups born /recipients	N chimaeras	>90% chimaerism	Survived >90% chimaerism
ntES2	diploid	217	12 (4)	14/4	10	5	3
ntES2	tetraploid	165	14 (6)	5/3	5	5	4
ntES3	diploid	250	13 (8)	12/4	12	3	1
ntES8	diploid	90	6 (3)	9/3	9	8	7
ntES8	tetraploid	180	15 (6)	7/3	7	7	6
ntES9	tetraploid	137	9 (5)	5/3	5	5	4
Control ES cells	tetraploid	313	22 (11)	19/6	19	19	4

Extended Data Table 3 | *In vivo* development of embryos derived by nuclear transfer into enucleated I2C or MII oocytes

Nuclear donor cells	Recipient cytoplasm	N embryos transferred/recipients	N pregnant recipients	Live pups born (%)
FFs	Interphase 2-cell embryo	115/10	4	0
	MII oocyte	170/10	4	1 (0.6%)
ES cells	Interphase 2-cell embryo	89/8	3	4 (4.5%)

doi:10.1038/nature13148

Quantitative proteomics identifies NCOA4 as the cargo receptor mediating ferritinophagy

Joseph D. Mancias[1,2,3,4], Xiaoxu Wang[1], Steven P. Gygi[2], J. Wade Harper[2] & Alec C. Kimmelman[1]

Autophagy, the process by which proteins and organelles are sequestered in double-membrane structures called autophagosomes and delivered to lysosomes for degradation, is critical in diseases such as cancer and neurodegeneration[1,2]. Much of our understanding of this process has emerged from analysis of bulk cytoplasmic autophagy, but our understanding of how specific cargo, including organelles, proteins or intracellular pathogens, are targeted for selective autophagy is limited[3]. Here we use quantitative proteomics to identify a cohort of novel and known autophagosome-enriched proteins in human cells, including cargo receptors. Like known cargo receptors, nuclear receptor coactivator 4 (NCOA4) was highly enriched in autophagosomes, and associated with ATG8 proteins that recruit cargo–receptor complexes into autophagosomes. Unbiased identification of NCOA4-associated proteins revealed ferritin heavy and light chains, components of an iron-filled cage structure that protects cells from reactive iron species[4] but is degraded via autophagy to release iron[5,6] through an unknown mechanism. We found that delivery of ferritin to lysosomes required NCOA4, and an inability of NCOA4-deficient cells to degrade ferritin led to decreased bioavailable intracellular iron. This work identifies NCOA4 as a selective cargo receptor for autophagic turnover of ferritin (ferritinophagy), which is critical for iron homeostasis, and provides a resource for further dissection of autophagosomal cargo–receptor connectivity.

Autophagosomes are decorated by a family of ubiquitin-like adaptor ATG8 proteins that are conjugated to phosphatidylethanolamine through the action of an autophagy-specific E1–E2–E3 cascade. Although ATG8 proteins are known to recruit a small number of cargo receptors to insipient autophagosomes, the full repertoire of selective autophagic cargo and their cognate receptor proteins remain poorly defined[3]. Selective autophagy may be particularly important for the survival or growth of particular cancer cell types[7,8] but in other contexts may act as a tumour suppressor to maintain normal cellular homeostasis and constrain tumour initiation[9,10]. Thus, a more comprehensive understanding of autophagy cargo–receptor pairs is required for understanding the autophagic mechanisms that contribute to proteostasis.

Three previous studies described the use of mass spectrometry to identify proteins in autophagosomal preparations, but the low overlap in the proteins identified between these studies and limitations of the approaches used led us to catalogue resident autophagosomal proteins using quantitative proteomics[11–13] (Extended Data Fig. 1a). We combined stable isotope labelling by amino acids in cell culture (SILAC) with an established density gradient separation protocol[14,15] to identify quantitatively proteins enriched in autophagosome preparations. This analysis was performed using two human pancreatic cancer cell lines (PANC-1 and PA-TU-8988T) that require autophagy for growth, as well as the MCF7 breast cancer cell line, which is less reliant on autophagy for growth[7]. Given the high basal autophagy of PANC-1 and PA-TU-8988T cells, light cells were briefly treated with the phosphoinositide 3-kinase inhibitor wortmannin to suppress autophagosome formation, whereas heavy cells were treated with the lysosomal inhibitor chloroquine to maximize the number of autophagosomes (Fig. 1a and Extended Data Fig. 1b). This approach allows for robust identification of proteins intimately associated with autophagosome-enriched samples as opposed to proteins that simply co-migrate with these vesicles during gradient centrifugation. As expected, in the autophagosome fraction the ATG8 protein MAP1LC3B (LC3B) was enriched, as assayed by immunoblotting or immunofluorescence, and contained characteristic double-membrane vesicles, as determined by electron microscopy (Extended Data Fig. 1c–h, k–m). These autophagosomes were intact, as assessed by LC3B and SQSTM1 (also known as p62) release upon detergent treatment (Extended Data Fig. 1i). We also note that autophagosomes and autophagolysosomes are heterogeneous in nature, as they form via a dynamic interplay between other membrane-rich organelles, each containing their own specific complement of proteins.

Single-label (heavy Lys) profiling of the autophagosomal fraction from PANC-1 cells after 4 or 16 h of chloroquine treatment, as well as double-label (heavy Lys and Arg) profiling of PANC-1- and MCF7-derived autophagosomal preparations at 16 h of chloroquine treatment resulted in the quantification of >2,000 proteins (Supplementary Tables 1–4; see Methods[16,17]). Proteins were selected on the basis of significantly increased $\log_2$(heavy:light) ratios and the presence of two or more peptides, and subsequently filtered against the relative abundance of the proteome, which was measured independently by liquid chromatography–mass spectrometry (LC–MS) (Fig. 1b; see Methods), thereby removing abundant proteins that may be non-specifically captured by bulk autophagy. We identified 86 proteins with $\log_2$(heavy:light) > 1.5 in all three PANC-1 replicates (Pearson correlation of 0.92 for a representative pair), and 102 proteins with $\log_2$(heavy:light) > 1.0 in both MCF7 replicates (Pearson correlation of 0.89) (Fig. 1c, d, Extended Data Fig. 1j and Supplementary Tables 3–5). We will refer to the union of these two high stringency data sets as class 1 autophagosome-enriched proteins, and a high priority subset of these proteins based on their presence in ≥3 data sets or known involvement in autophagy as class 1A (Fig. 1e, Extended Data Fig. 2a and Supplementary Table 5). We also identified non-class-1 proteins with $\log_2$(heavy:light) > 2.0 in any two of the five PANC-1 or MCF7 experiments (16 h in chloroquine), and will refer to this lower stringency data set as class 2 autophagosome-enriched proteins (Supplementary Table 5). As expected, PANC-1 cells treated with chloroquine for 16 h showed a greater accumulation of candidate proteins than PANC-1 or PA-TU-8988T cells treated for 4 h (Fig. 1e and Extended Data Fig. 2).

Within the class 1A proteins, we identified two ATG8 paralogues (GABARAPL2 and MAP1LC3B), four known autophagy cargo receptors (SQSTM1, CALCOCO2 (also known as NDP52), OPTN and NBR1) and four proteins that were previously reported to associate with ATG8 family members and/or cargo receptors, or to be involved in autophagosomal membrane fusion (KEAP1, TMEM59, FYCO1 and STX17) (Fig. 1e and Extended Data Fig. 2a). Moreover, we identified two proteins reported as selective autophagy cargo (APP and NRP1), and seven proteins identified as high-confidence candidate interacting proteins (HCIPs) in our

[1]Division of Genomic Stability and DNA Repair, Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [2]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Harvard Radiation Oncology Program, Boston, Massachusetts 02115, USA. [4]Department of Radiation Oncology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA.

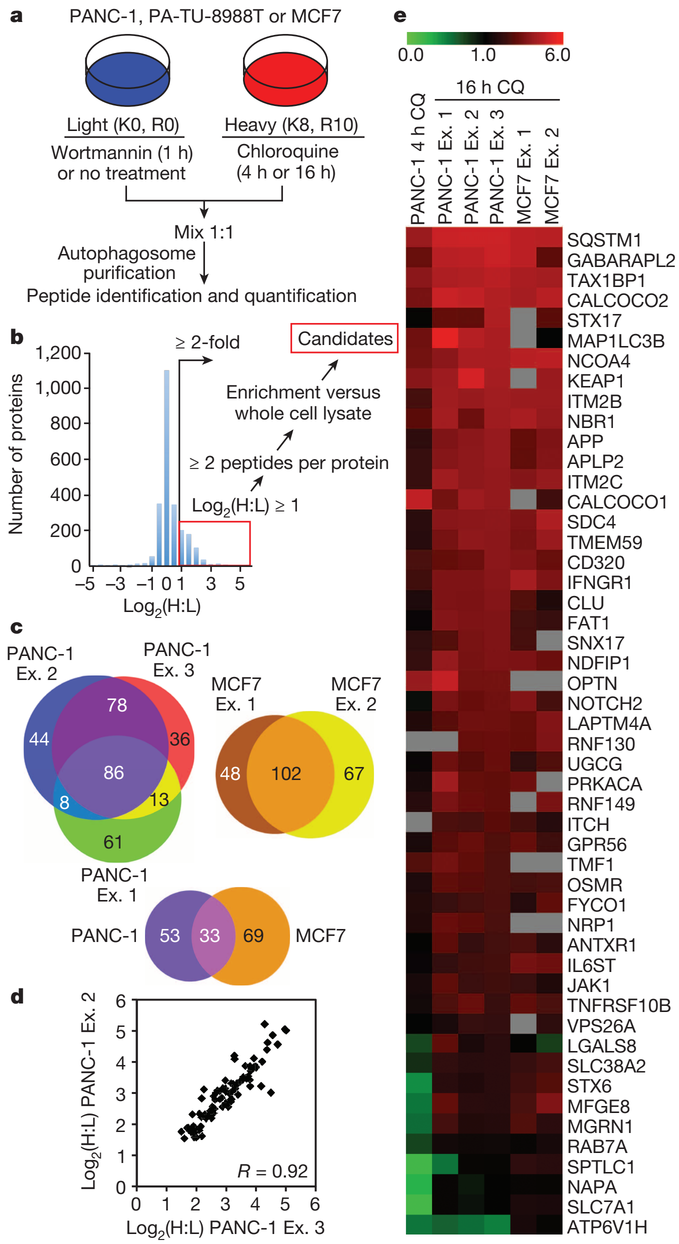


Figure 1 | Quantitative proteomics for identification of autophagosome-associated proteins. **a**, Autophagosome enrichment workflow. **b**, Log_2(heavy (H):light (L)) plot for autophagosome proteins from PANC-1 cells (experiment (Ex.) 3, Supplementary Table 3) and scheme for identification of candidate autophagosome proteins. **c**, Autophagosome candidate overlap from biological replicate experiments for PANC-1 and MCF7 cells, as well as overlap between PANC-1 and MCF7 data sets. **d**, Pearson correlation plot for overlapping candidates from PANC-1 experiments (86 proteins, comparing experiment 2 with experiment 3). *x* axis: log_2(H:L) PANC-1 experiment 3 (L-wortmannin:H-chloroquine 16 h) value; *y* axis: log_2(H:L) PANC-1 experiment 2 (L-wortmannin: H-chloroquine 16 h) value. **e**, Log_2(H:L) heat map of class 1A candidates from PANC-1 and MCF7 cells. CQ, chloroquine.

previously published autophagy interaction network[18] (Fig. 1e and Extended Data Fig. 2a). Furthermore, the autophagy pathway proteins RB1CC1, ATG9A and TBC1D15, and the OPTN-binding protein TBK1, were identified in the class 2 data set (Supplementary Table 5). The presence of several plasma membrane and endocytosis-related proteins (Fig. 1e and Extended Data Fig. 2a) is consistent with intermixing of these membrane sources during autophagosome maturation or lysosomal fusion. Co-localization with LC3B-positive puncta was observed for 7 of 11 class 1A proteins tested (Extended Data Figs 2a and 3).

As further validation, PA-TU-8988T pancreatic cancer cells were subjected to both SILAC-based autophagosomal profiling using density gradient purification and a semi-quantitative proteomic approach using an immunopurification-based autophagosome enrichment scheme by immunoprecipitation of green fluorescent protein (GFP)-tagged LC3B autophagosomes[12] (Extended Data Figs 2b–d, 4 and Supplementary Tables 6, 7). In total, 40 proteins were identified in common between the 50 class 1A autophagosomal proteins and those identified as enriched in autophagosomes purified from PA-TU-8988T cells. Comparative analysis of our MCF7 candidate proteins with a previously reported autophagosomal quantitative proteomics analysis[13] that used the same cell line revealed two overlapping proteins from their final combined data set (SQSTM1 and GABARAPL2) (Extended Data Fig. 5 and Supplementary Table 8). Expanding the analysis to also include our PANC-1 data sets only increased the overlap to a total of four proteins from our class 1 and 2 lists. This low rate of overlap seems to reflect the fact that a large fraction of proteins previously reported to reside in autophagosomes would have been removed upon filtering at high stringency for abundance in the total proteome (see Methods; Extended Data Fig. 5 and Supplementary Table 8), indicating that these are probably present in autophagosomes owing to non-selective bulk degradation of cytosolic contents, or are co-purifying contaminants. Likewise there is minimal overlap between our class 1 and 2 proteins with two additional proteomics efforts[11,12]. Although it is understandable that there would be variation between data sets depending on cell type, autophagy stimulus and purification technique, our data set represents the most robust autophagosome proteomics effort so far given the number of bona fide autophagy-related proteins identified among our class 1 and 2 candidates.

Among the most highly and consistently enriched autophagosomal proteins was NCOA4 (Fig. 1e and Supplementary Tables 2–4, 6 and 7), which was also enriched in an independent autophagosome proteomics study, although not included in their final list of autophagosomal proteins owing to the method of analysis[13] (Extended Data Fig. 5 and Methods). NCOA4 was originally identified as a protein that interacts with the androgen receptor (AR), and its overexpression was reported to activate transcription of AR-regulated genes[19]. However, not all studies have supported a role for NCOA4 in AR function[20], and the data described later reveals a previously unrecognized role for NCOA4 as an autophagy cargo receptor.

We initially examined the localization of NCOA4. GFP–NCOA4 was diffusely localized in the cytoplasm in U2OS human osteosarcoma cells that have a low level of basal autophagy but accumulated in cytoplasmic puncta that are largely co-incident with LC3B-positive puncta in response to chloroquine in U2OS cells as well as PA-TU-8988T cells, consistent with localization of GFP–NCOA4 in autophagosomes (Fig. 2a and Extended Data Fig. 6a). In addition, NCOA4 showed significant co-localization with GABARAPL2-positive puncta (Extended Data Fig. 6b, c), which is consistent with it being among the strongest interactors in a glutathione *S*-transferase (GST)–ATG8 binding assay performed *in vitro* using cell extracts (Fig. 2b). Consistent with our proteomic data, NCOA4 is enriched in purified autophagosomes by immunoblotting and its levels are markedly increased in response to blockade of autophagosome degradation by chloroquine or bafilomycin A1 (BAF) (Fig. 2c, d), Moreover, NCOA4 does not co-localize with the late endosome marker mannose 6-phosphate receptor (Extended Data Fig. 6d). Together these data support the predominantly autophagosomal localization of NCOA4 puncta. We were unable to identify a canonical LC3-interacting region (LIR) motif in NCOA4, although the existence of non-canonical ATG8-binding motifs[21] may suggest that NCOA4 uses such an alternative motif for interaction.

To begin to understand potential roles for NCOA4 in autophagy, we performed affinity purification–mass spectrometry (AP–MS) of stably expressed haemagglutinin (HA)- and Flag-tagged NCOA4 and used the Comparative Proteomics Software Analysis Suite (CompPASS) to identify HCIPs[22]. AP–MS of NCOA4–HA–Flag from PANC-1, PA-TU-8988T

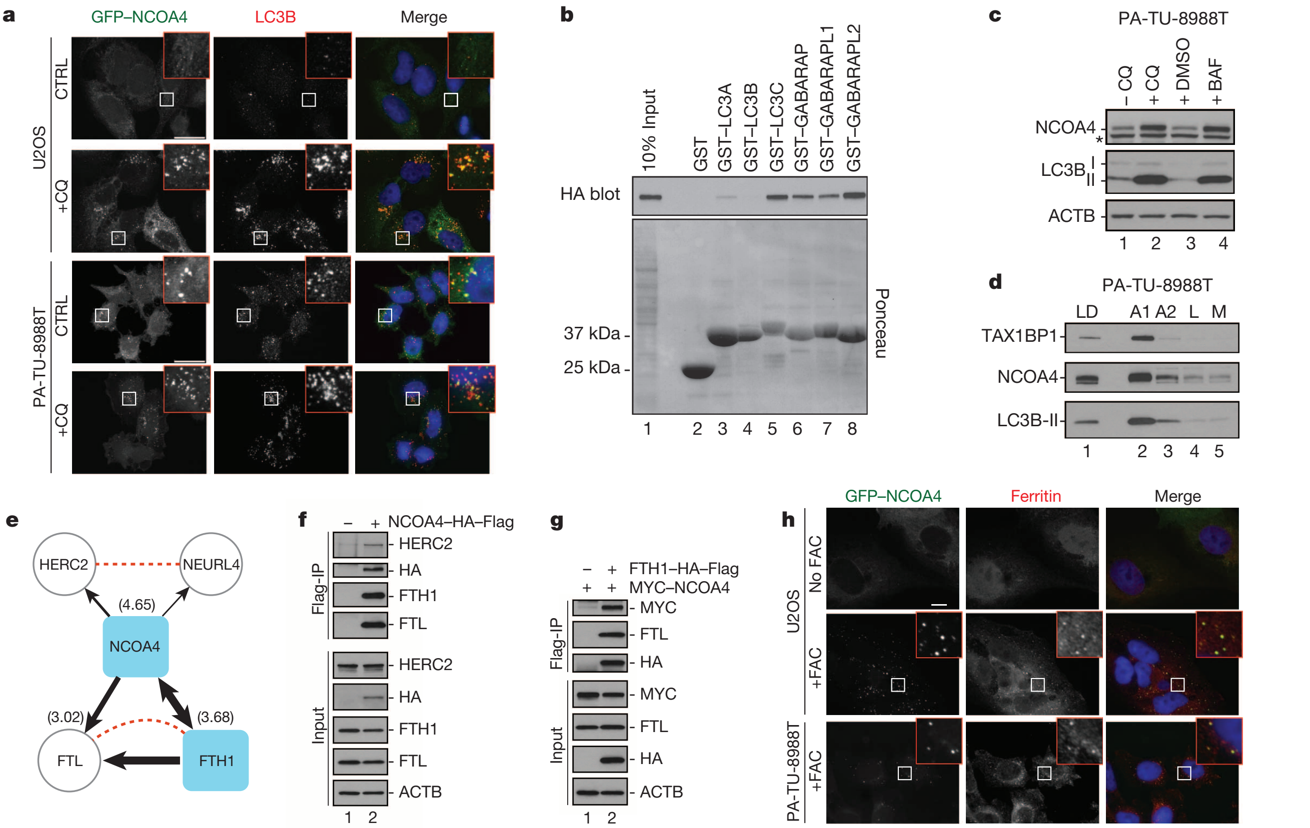


Figure 2 | NCOA4 associates with and co-localizes with ferritin. **a**, GFP–NCOA4 (green) co-localizes with endogenous LC3B (red) in chloroquine (CQ)-treated cells. CTRL, control. Scale bar, 20 μm. **b**, GST pull-down assay of NCOA4–HA–Flag from stable 293T cells using GST–ATG8 proteins. HA immunoblot for NCOA4–HA–Flag. **c**, Lysates from PA-TU-8988T cells treated with chloroquine or BAF (8 h) were immunoblotted for NCOA4, LC3B and ACTB as a loading control. Asterisk indicates cross-reactive band (see Extended Data Fig. 7d, e). **d**, PA-TU-8988T autophagosome purification fractions were analysed using antibodies to NCOA4 and LC3B. TAX1BP1, a newly identified autophagy receptor, was included as a positive control. A1, autophagosome fraction; A2, autophagolysosome fraction; L, lysosome fraction; LD, gradient load; M, mitochondrial fraction. **e**, NCOA4 interaction network from cells expressing NCOA4–HA–Flag or FTH1–HA–Flag (Supplementary Table 9). Black lines (this study) depict directionality of interaction observed with line thickness weighted by weighted and normalized *D* score (WDN) score (293T data set). Dotted lines indicate data derived from the STRING database. Numbers in parentheses indicate $\log_2$(H:L) ratio of NCOA4, FTH1 and FTL from MCF7 Ex. 1 data set (Supplementary Table 4). **f**, **g**, Extracts from 293T cells stably expressing the indicated proteins were immunoprecipitated with anti-Flag (Flag-IP) and immunoblotted with the indicated antibodies. Anti-ACTB, loading control. **h**, Representative confocal images of GFP–NCOA4 (green) and ferritin (red) after no treatment or FAC treatment (24 h). Scale bar, 10 μm.

and 293T cells revealed a number of HCIPs, including both the ferritin heavy chain (FTH1) and ferritin light chain (FTL), as well as HERC2 and NEURL4, which are known to associate with each other[23] (Fig. 2e and Supplementary Table 9). Interaction of NCOA4–HA–Flag with selected endogenous HCIPs was verified by anti-Flag immunoprecipitation followed by immunoblotting (Fig. 2f). FTH1–HA–Flag reciprocally associated with endogenous NCOA4 as determined by AP–MS (Fig. 2e and Supplementary Table 9) and also associated with co-expressed MYC-tagged NCOA4 (Fig. 2g). The absence of HERC2 and NEURL4 in ferritin immune complexes (Supplementary Table 9) suggests that NCOA4 makes distinct complexes with ferritin and HERC2–NEURL4. Consistent with this, neither HERC2 nor NEURL4 were enriched in autophagosomal fractions (Supplementary Tables 2–7) and HERC2 does not co-localize with autophagosomes (Extended Data Fig. 6e).

FTH1 and FTL form a 24-subunit macromolecular iron-storage complex critical for iron homeostasis[4]. Early electron microscopy studies identified iron-laden ferritin in lysosomes and, more recently, it was shown that in cells subjected to iron chelation, ferritin is delivered to the lysosome for degradation via autophagy, presumably to promote iron availability[5,6,24]. Interestingly, FTH1 and FTL were enriched in autophagosomal fractions from MCF7 and PA-TU-8988T cells, as determined by mass spectrometry (Fig. 2e, Extended Data Fig. 2 and Supplementary Tables 4 and 6). Furthermore, ferritin and NCOA4 demonstrated extensive co-localization in puncta in several cell lines upon stimulation of ferritin expression with ferric ammonium citrate (FAC), reflecting the high level of ferritin undergoing autophagic targeting and lysosomal degradation (Fig. 2h). We refer to this process as 'ferritinophagy'.

Given the interaction between NCOA4 and ferritin, and their localization in autophagosomes, we tested the hypothesis that NCOA4 acts as an autophagy receptor for ferritinophagy. In response to low intracellular iron levels, ferritin is degraded to release its iron stores. This can be stimulated experimentally by chelation of iron[5]. Although some reports have suggested that ferritin is degraded via the proteasome[25], we validated that in the cell lines we examined, ferritin is primarily degraded by the lysosome in response to multiple distinct chelators (Extended Data Fig. 7a, b), as observed previously[5]. In addition, genetic inhibition of autophagy using RNA interference (RNAi) against ATG5 reduced ferritin degradation in response to iron chelation (Extended Data Fig. 7c). Importantly, suppression of NCOA4 expression with multiple short hairpin RNAs (shRNAs) followed by iron chelation abrogated ferritin degradation in multiple cell lines and with chemically distinct chelators (Fig. 3a, b and Extended Data Fig. 7d–g). Consistent with these results, ferritin localized to lysosomes and accumulated to a significant degree upon lysosomal protease inhibition (Extended Data Fig. 8a, b). Furthermore, ferritin simultaneously co-localizes with NCOA4- and LC3B-positive puncta representing autophagosomes (Fig. 3c). If NCOA4 functions as

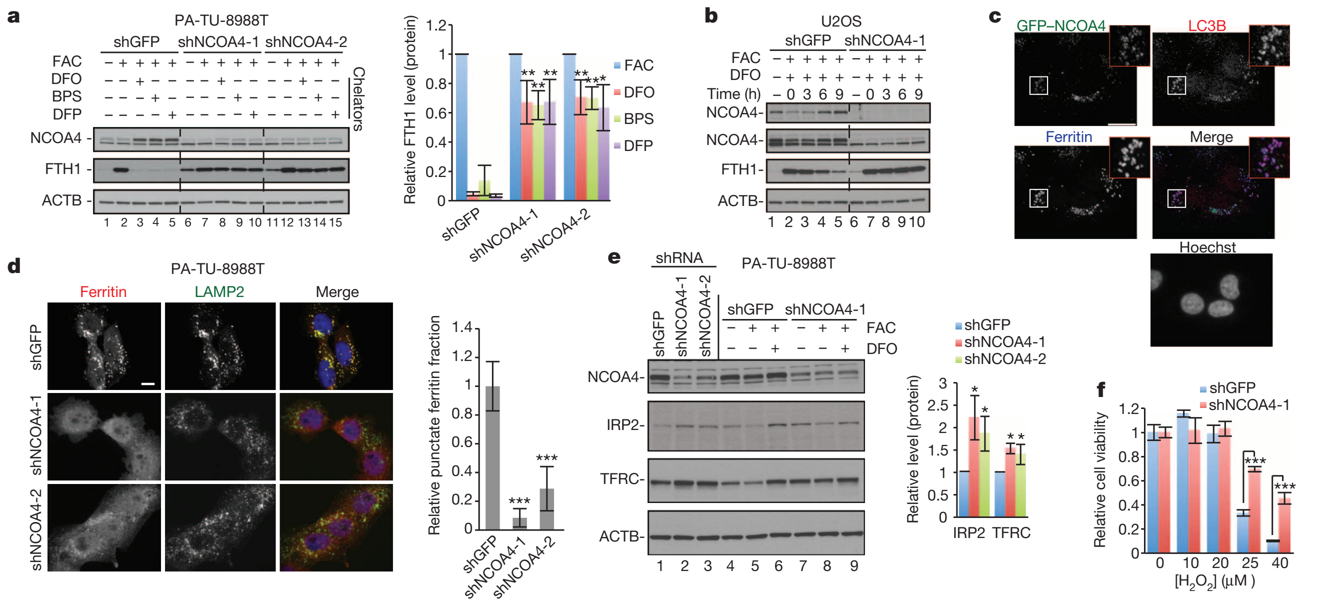


Figure 3 | NCOA4 mediates autophagic delivery of ferritin to lysosomes to control iron homeostasis. **a**, NCOA4 depletion rescues ferritin degradation upon 9 h of iron chelation in PA-TU-8988T cells. Relative FTH1 levels ($n = 3$, biological triplicate) for each chelator are quantified. Bars and error bars represent mean values and standard deviation (s.d.), respectively. $^{**}P < 0.01$ and $^{*}P < 0.02$ comparing FTH1 levels between different cell lines (one-sided t-test). **b**, Deferoxamine mesylate (DFO) chelation time course in U2OS cells. DFO added at time 0, two NCOA4 antibodies are used for immunoblotting (top panel, Sigma antibody; second panel, Bethyl antibody). **c**, GFP–NCOA4 (green) co-localizes with endogenous LC3B (red) and endogenous ferritin (blue) in U2OS cells subjected to DFO chelation in the presence of lysosomal protease inhibitors for 6 h. Scale bar, 20 μm. **d**, Immunostaining of PA-TU-8988T cells subjected to DFO chelation in the presence of lysosomal protease inhibitors for 9 h. Scale bar, 10 μm. Punctate ferritin fraction was quantified from >100 cells per cell line from two independent experiments (biological duplicate). Bars and error bars represent mean values and s.d., respectively. $^{***}P < 0.001$ using a one-sided t-test. **e**, Lysates from PA-TU-8988T cells as in **a** were analysed using indicated antibodies (lanes 1–3, untreated), with quantification based on at least three independent experiments (biological triplicate). Bars and error bars represent mean values and s.d., respectively. $^{*}P < 0.05$ using a one-sided t-test. **f**, PA-TU-8988T cells stably expressing shGFP or shNCOA4-1 were treated with H_2O_2 and cell viability was measured at 72 h. Bars and error bars represent mean values and s.d., respectively, of technical triplicates. $^{***}P < 0.001$ using a two-sided t-test.

an essential receptor for ferritinophagy, depletion of NCOA4 would be predicted to block ferritin localization with lysosomes. Indeed, depletion of NCOA4 blocked ferritin co-localization with lysosomes and led to a diffuse localization pattern (Fig. 3d and Extended Data Fig. 8c). NCOA4 is also critical in mediating the targeting of ferritin to autophagosomes in non-transformed IMR-90 human fibroblasts and human pancreatic duct epithelial (HPDE) cells, indicating that this is a general cellular mechanism for the degradation of ferritin by autophagy (Extended Data Fig. 8d, e). As a control for the specificity of the shRNAs, a non-degradable murine *Ncoa4* complementary DNA (cDNA) was able to rescue significantly the ferritin lysosomal localization (Extended Data Fig. 9a–c). These results were confirmed with two additional short interfering RNAs (siRNAs) against NCOA4 (Extended Data Fig. 10a). Again, consistent with a distinct role for the HERC2–NCOA4 complex, HERC2 knockdown had no effect on lysosomal delivery of ferritin (Extended Data Fig. 10b, c).

Iron has an essential role in multiple cellular processes and, as such, iron metabolism is a tightly regulated process controlled by a network of iron-dependent proteins[4]. Ferritin forms an intracellular iron storage protein complex capable of chelating up to 4,500 atoms of iron, thus protecting the cell from free iron participating in the generation of free radicals via Fenton-like reactions[26]. Iron stored in ferritin is used during periods of low iron levels and recent evidence suggests that iron can be liberated from ferritin upon degradation of ferritin in the lysosome[5]. As our evidence suggested that NCOA4 mediates the transport of ferritin to the lysosome via the autophagosome, we were interested in how NCOA4 knockdown affects the network of iron regulatory proteins. First, NCOA4 depletion increased basal ferritin levels relative to control cells (Fig. 3a, lane 1 versus lanes 6 and 11, Extended Data Fig. 7f, lane 1 versus lanes 5 and 9). This suggests that NCOA4-deficient cells would have reduced iron bioavailability. Therefore, we examined the levels of iron-responsive-element-binding protein 2 (IRP2), an RNA-binding protein that binds to iron-responsive elements (IREs) on a set of iron-regulated messenger RNAs to control their translation[4]. Cellular iron levels control IRE-binding activity of IRP2, with high and low iron levels promoting IRP2 turnover and stabilization, respectively[4]. We found that NCOA4 depletion increased IRP2 abundance to an extent comparable with control cells subjected to iron chelation (Fig. 3e, lane 7 versus lane 6). A higher IRP2 level in NCOA4-depleted cells would be anticipated to stabilize transferrin receptor mRNA and increase translation of the transferrin receptor to increase intracellular iron. Indeed, the abundance of the transferrin receptor is increased in NCOA4-depleted PA-TU-8988T cells (Fig. 3e, lane 1 versus lanes 2 and 3). These results were confirmed with two independent siRNAs to NCOA4 in PA-TU-8988T, U2OS and IMR-90 cell lines (Extended Data Fig. 10d). Conversely, ectopic expression of NCOA4 in PA-TU-8988T cells attenuates the increase in ferritin seen in FAC-treated control cells, implying that the increased NCOA4 is promoting ferritin turnover (Extended Data Fig. 10e). We also note that the abundance of NCOA4 was altered by iron loading or chelation, although no canonical IRE was identified[4] (Fig. 3b and Extended Data Fig. 10e).

We next examined the biological consequences of reduced iron availability and lysosomal ferritin degradation in response to NCOA4 depletion. Lysosomal iron has been shown to react with reactive oxygen species (ROS), leading to free radical formation via Fenton-like reactions, which results in lysosomal bursting and cell death[26]. Therefore, NCOA4 knockdown should provide protection from cell death after ROS challenge. As shown in Fig. 3f, control cells were more sensitive to hydrogen peroxide challenge than NCOA4-depleted cells.

Selective autophagy is increasingly recognized as a regulated process through which specific cellular proteins, complexes and organelles are degraded in the lysosome in response to diverse stimuli. Previous work

suggested that delivery of the ferritin complex to lysosomes occurs via autophagy and regulates iron bioavailability[5,6]. Our identification of NCOA4 as a specific cargo receptor for ferritin provides the first mechanistic understanding of how the ferritin complex is selectively delivered to autophagosomes. Flux through the pathway is regulated by iron availability and, in turn, disruption of the pathway through modulation of NCOA4 levels leads to alterations in the activity of IRP2 and altered sensitivity of cells to ROS. Moreover, *NCOA4* mRNA is induced in red blood cells during erythropoiesis, and its expression correlates with genes involved in haem biosynthesis[27], raising the possibility that NCOA4 function is important for both cellular remodelling and iron availability during differentiation[28]. Although our data provide compelling evidence for the role of NCOA4 as a ferritinophagy cargo receptor, we cannot rule out it having other roles in specific cellular contexts. However, given our findings in normal and tumour cells of diverse tissue origin, our work suggests that the targeting of ferritin to autophagosomes by NCOA4 is a general cellular mechanism for regulating bioavailable iron. Although further functional studies are required to elucidate whether selective autophagy of particular cargo underlies the strong reliance of pancreatic and other cancer cells on autophagy for growth, our work reveals the potential of quantitative autophagy proteomics to uncover receptor–cargo relationships and to elucidate further the mechanisms underlying both macro and selective autophagy.

METHODS SUMMARY

Cells were grown in Lys(K)-free and Arg(R)-free DMEM/dialysed fetal bovine serum, with light (K0, R0) or heavy (K8/R0 or K8/R10) Lys/Arg, and treated with chloroquine (25 μM) for 4 or 16 h. Mixed cells were lysed and autophagosomes were purified as described[14,15], before SDS–PAGE and in-gel digestion with trypsin or Lys-C, and LC–MS/MS. Candidate autophagosomal proteins were identified by using a multi-step filtering process including a $\log_2$(heavy:light) ratio cut-off and enrichment in autophagosomes versus whole proteome. Interaction proteomics used 293T, PA-TU-8988T or PANC-1 cells stably expressing NCOA4 or FTH1 HA–Flag-tagged constructs and CompPASS to identify HCIPs[22]. Chelation assays were performed similarly to those previously described[5] with slight modifications. Cells stably expressing shRNAs to NCOA4 were cultured for 24 h in iron-enriched media (supplementary FAC added) and subsequently subjected to iron chelation with multiple different chelators before immunoblotting or immunofluorescence.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 September 2013; accepted 12 February 2014.
Published online 30 March 2014.

1. Yang, Z. & Klionsky, D. J. Eaten alive: a history of macroautophagy. *Nature Cell Biol.* **12,** 814–822 (2010).
2. Kroemer, G., Marino, G. & Levine, B. Autophagy and the integrated stress response. *Mol. Cell* **40,** 280–293 (2010).
3. Kirkin, V., McEwan, D. G., Novak, I. & Dikic, I. A role for ubiquitin in selective autophagy. *Mol. Cell* **34,** 259–269 (2009).
4. Pantopoulos, K., Porwal, S. K., Tartakoff, A. & Devireddy, L. Mechanisms of mammalian iron homeostasis. *Biochemistry* **51,** 5705–5724 (2012).
5. Asano, T. *et al.* Distinct mechanisms of ferritin delivery to lysosomes in iron-depleted and iron-replete cells. *Mol. Cell. Biol.* **31,** 2040–2052 (2011).
6. Kidane, T. Z., Sauble, E. & Linder, M. C. Release of iron from ferritin requires lysosomal activity. *Am. J. Physiol. Cell Physiol.* **291,** C445–C455 (2006).
7. Yang, S. *et al.* Pancreatic cancers require autophagy for tumor growth. *Genes Dev.* **25,** 717–729 (2011).
8. Sandilands, E. *et al.* Autophagic targeting of Src promotes cancer cell survival following reduced FAK signalling. *Nature Cell Biol.* **14,** 51–60 (2012).
9. Kimmelman, A. C. The dynamic nature of autophagy in cancer. *Genes Dev.* **25,** 1999–2010 (2011).
10. White, E. Deconvoluting the context-dependent role for autophagy in cancer. *Nature Rev. Cancer* **12,** 401–410 (2012).
11. Øverbye, A., Fengsrud, M. & Seglen, P. O. Proteomic analysis of membrane-associated proteins from rat liver autophagosomes. *Autophagy* **3,** 300–322 (2007).
12. Gao, W. *et al.* Biochemical isolation and characterization of the tubulovesicular LC3-positive autophagosomal compartment. *J. Biol. Chem.* **285,** 1371–1383 (2010).
13. Dengjel, J. *et al.* Identification of autophagosome-associated proteins and regulators by quantitative proteomic analysis and genetic screens. *Mol. Cell. Proteomics* **11,** M111.014035 (2012).
14. Marzella, L., Ahlberg, J. & Glaumann, H. Isolation of autophagic vacuoles from rat liver: morphological and biochemical characterization. *J. Cell Biol.* **93,** 144–154 (1982).
15. Koga, H., Kaushik, S. & Cuervo, A. M. Altered lipid content inhibits autophagic vesicular fusion. *FASEB J.* **24,** 3052–3065 (2010).
16. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143,** 1174–1189 (2010).
17. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **1,** 2856–2860 (2007).
18. Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. Network organization of the human autophagy system. *Nature* **466,** 68–76 (2010).
19. Yeh, S. & Chang, C. Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. *Proc. Natl Acad. Sci. USA* **93,** 5517–5521 (1996).
20. Gao, T., Brantley, K., Bolu, E. & McPhaul, M. J. RFG (ARA70, ELE1) interacts with the human androgen receptor in a ligand-dependent fashion, but functions only weakly as a coactivator in cotransfection assays. *Mol. Endocrinol.* **13,** 1645–1656 (1999).
21. von Muhlinen, N. *et al.* LC3C, bound selectively by a noncanonical LIR motif in NDP52, is required for antibacterial autophagy. *Mol. Cell* **48,** 329–342 (2012).
22. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138,** 389–403 (2009).
23. Martínez-Noël, G. *et al.* Identification and proteomic analysis of distinct UBE3A/E6AP protein complexes. *Mol. Cell. Biol.* **32,** 3095–3106 (2012).
24. Trump, B. F., Valigorsky, J. M., Arstila, A. U., Mergner, W. J. & Kinney, T. D. The relationship of intracellular pathways of iron metabolism to cellular iron overload and the iron storage diseases. Cell sap and cytocavitary network pathways in relation to lysosomal storage and turnover of iron macromolecules. *Am. J. Pathol.* **72,** 295–336 (1973).
25. De Domenico, I. *et al.* Ferroportin-mediated mobilization of ferritin iron precedes ferritin degradation by the proteasome. *EMBO J.* **25,** 5396–5404 (2006).
26. Kurz, T., Gustafsson, B. & Brunk, U. T. Intralysosomal iron chelation protects against oxidative stress-induced cellular damage. *FEBS J.* **273,** 3106–3117 (2006).
27. Nilsson, R. *et al.* Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metab.* **10,** 119–130 (2009).
28. Griffiths, R. E. *et al.* The ins and outs of human reticulocyte maturation: autophagy and the endosome/exosome pathway. *Autophagy* **8,** 1150–1151 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Everley and E. Huttlin for assistance with LC–MS/MS, A. Cuervo for assistance with autophagosome purification protocol development, Y. Liu for technical support with cell culture, A. White, J. Lydeard and S. Hayes for interaction proteomics support, M. Haigis for access to nitrogen cavitation equipment, and the Nikon Imaging and Electron Microscopy Centers (Harvard Medical School) for imaging support. We thank Millennium Pharmaceuticals for the GABARAPL2 antibody. We acknowledge support from National Institutes of Health grants GM070565 and GM095567 to J.W.H., and National Cancer Institute grant R01CA157490, American Cancer Society Research Scholar grant RSG-13-298-01-TBG and the Lustgarten Foundation to A.C.K. J.D.M. was supported by an American Board of Radiology Holman Research Pathway Post-doctoral Fellowship and an American Society of Radiation Oncology Junior Faculty Career Research Training Award (JF2013-2). J.W.H. is a consultant for Millennium: The Takeda Oncology Company and Biogen Idec. A.C.K. is a consultant for Forma Therapeutics.

Author Contributions J.D.M., J.W.H. and A.C.K. conceived the experiments. J.D.M. performed all experiments. X.W. assisted with cell line generation. S.P.G. provided proteomics software and analysis support. J.D.M., J.W.H. and A.C.K. analysed data and wrote the manuscript. All authors edited the manuscript.

Author Information Processed proteomics data are available in Supplementary Information; raw data files have been deposited at Peptide Atlas under accession number PASS00440. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C.K. (alec_kimmelman@dfci.harvard.edu) or J.W.H. (wade_harper@hms.harvard.edu).

METHODS

Cell culture and reagents. PANC-1, PA-TU-8988T, MCF7, U2OS, IMR-90 and 293T cell lines were obtained from the American Type Culture Collection or the German Collection of Microorganisms and Cell Cultures. HPDE cells were cultured as previously described[7]. All cell lines were tested routinely for mycoplasma contamination.

Antibodies. The following antibodies were used in these studies. LAMP2 (Abcam Ab25631; western blot 1:1.000; immunofluorescence 1:100); HA (Covance MMS 101P; western blot 1:2,000; immunofluorescence 1:100); LC3B (Cell Signaling 2775; western blot 1:2,000); LC3B (Cell Signaling 3868; immunofluorescence 1:200); LC3B (nanoTools 0231-100/LC3-5F10; immunofluorescence 1:100); VDAC1 (Abcam Ab28777; western blot 1:1,000); ferritin (Rockland 200-401-090-0100; Immunofluorescence 1:400); FTH1 (Cell Signaling 3998; western blot 1:1,000); NCOA4 (ARA70) (Bethyl Laboratories A302-272A; western blot 1:1,000); NCOA4 (Sigma SAB1404569; western blot 1:1,000); SQSTM1 (Abnova H00008878-M01; western blot 1:5,000); TAX1BP1 (Cell Signaling 8182; western blot 1:1,000); ATG5 (Cell Signaling 2630; western blot 1:1,000); HERC2 (BD Transduction Laboratories 612366; western blot 1:1,000); FTL (Abnova Ab69090; western blot 1:1,000); ACTB (Sigma A2066; western blot 1:5,000); MYC (Santa Cruz Sc-40; western blot 1:1,000); IRP2 (Santa Cruz Sc-33682; western blot 1:500); TFRC (BD Transduction Laboratories 612124; western blot 1:2,000); mannose 6-phosphate receptor (Abcam Ab2733; Immunofluorescence 1:100); GABARAPL2 (gift from Millennium Pharmaceuticals; Immunofluorescence 1:100). The following secondary antibodies were used. Anti-rabbit IgG (H+L) HRP conjugate (western blot secondary 1:7,500); anti-mouse IgG (H+L) HRP conjugate (western blot secondary 1:7,500); Alexa Fluor 488 anti-mouse IgG (H+L) (Immunofluorescence secondary 1:1,000); Alexa Fluor 594 anti-rabbit IgG (H+L) (Immunofluorescence secondary 1:1,000), Alexa Fluor 633 anti-rabbit IgG (H+L) (Immunofluorescence secondary 1:1,000).

RNAi. *ATG5* siRNA (NM_004849.3) was purchased from Invitrogen as previously published[7]. siRNAs were transfected using a reverse transfection protocol and RNAiMax (Invitrogen). siControl was an siRNA designed against luciferase. Lentiviral shRNA plasmid clones (pLKO.1) were obtained from the RNAi Consortium collection. shNCOA4-1: 5′-CCCAGGAAGTATTACTTAATT-3′ (TRCN0000019724); shNCOA4-2: 5′-GCTGGCAAACAGAAGTTTAAA-3′ (TRCN0000019726); and shGFP: 5′-GCAAGCTGACCCTGAAGTTCAT-3′ (Addgene plasmid #30323) (*NCOA4* accession number NM_001145263.1). Lentivirus was produced as described previously[7]. Additional siRNAs used in this work were purchased from Invitrogen against *NCOA4* (NM_001145263.1); siNCOA4-1: 5′-ACAAAGAUCUAGCCAAUCA-3′; and siNCOA4-2: 5′-GACCUUAUUUAUCAGCUUA-3′; and against *HERC2* (NM_004667.5); siHERC2-1: 5′-GCACAGAGUAUCACAGGUA-3′; and siHERC2-2: 5′-CGAUGAAGGUUUGGUAUUU-3′.

Chemicals. FAC (Fisher Scientific, 36–180 μM titrated for each cell line based on level of FTH1 translation at 24 h after addition of FAC), DFO (BioVision; 100 μM), bathophenanthroline disulphonate (BPS; Sigma; 300 μM), deferiprone (DFP; Sigma; 100 μM), deferasirox (DFX; Selleckchem; 30 μM), E64-d (Sigma; 10 μg ml^{-1}), pepstatin A (CalBiochem; 10 μg ml^{-1}), bortezomib (BTZ; gift from Millenium Pharmaceuticals; 1 μM), chloroquine (Sigma; 10 or 25 μM), BAF (Sigma; 50 nM), hydrogen peroxide (H_2O_2; Sigma), CellTiter-Glo (Promega).

SILAC-based density gradient centrifugation autophagosome enrichment. PANC-1, PA-TU-8988T and MCF7 cells were grown in Lys- and Arg-free DMEM supplemented with 10% dialysed FBS (Gibco), 2 mM L-glutamine, penicillin-streptomycin, and light (K0) Lys (50 μg ml^{-1}) and light (R0) Arg (85 μg ml^{-1}). Heavy media was the same except the light Lys was replaced with K8 Lys (Cambridge Isotopes) and the light Arg was replaced with R10 Arg at the same concentrations (PANC-1 and PA-TU-8988T cell lines were also grown in K8-only heavy media (K8, R0)). Where indicated, cells (10^8) were treated with wortmannin (200 nM) or chloroquine (25 μM) for the times indicated. After the indicated treatments, heavy and light cells were mixed 1:1 by cell number. Autophagosome purification was performed as described previously[14,15] with slight modifications. All steps were carried out at 4 °C. Briefly, cells were washed three times with PBS and resuspended in buffer A (250 mM sucrose, 10 mM HEPES, pH 7.4, 1 mM EDTA, protease inhibitors (EDTA-free, Roche)). Cells were lysed by nitrogen cavitation and homogenized using a potter-elvehjem homogenizer with a teflon pestle. Lysates were centrifuged at 2,000*g*, the supernatant was centrifuged at 17,000*g* and resuspended in 0.95 ml buffer A. Lysate was diluted with 1.45 ml 85.6% Nycodenz (Sigma-Aldrich) solution. A discontinuous Nycodenz gradient (26%, 24%, 20%, 15%) was layered on top of the lysate and spun at 24,700 r.p.m. in a SW41 rotor (Beckman). Fractions were collected as indicated, pelleted at 24,000*g*, and used for downstream applications (immunoblotting, electron microscopy, mass spectrometry).

Autophagosome immunoisolation. PA-TU-8988T cells stably expressing GFP–LC3B were treated as described earlier. Lysis and clarification centrifugation steps are as described earlier. GFP-based immunoisolation was performed as previously described[12]. Briefly, lysates were incubated with μMACS microbeads (magnetic beads coated with anti-GFP; MACS Miltenyi Biotec) for 1 h at 4 °C with mixing. Lysate–bead mixture was applied to a LS Column in a MidiMACS Separator, washed, eluted, pelleted and prepared for downstream analysis (mass spectrometry).

Mass spectrometry analysis of autophagosomes. Enriched autophagosomes were prepared for mass spectrometry analysis as described[17]. Briefly, pelleted autophagosomes were resuspended in 2% SDS, 50 mM Tris, pH 7.5, 2 mM EDTA, boiled for 10 min, and centrifuged at 16,100*g* for 5 min at room temperature. Supernatants were subjected to SDS–PAGE followed by in-gel digestion with trypsin (K8, R10 heavy samples and PA-TU-8988T K8 sample) or Lys-C (PANC-1 K8 heavy samples). For generation of comparison whole-cell lysate data sets, untreated heavy- and light-labelled cells were mixed in a 1:1 ratio, lysed as described earlier, and subjected to SDS–PAGE followed by in-gel digestion. Approximately 10 μg of total whole-cell lysate was used for this analysis, however, multiple dilutions of whole-cell lysate extract were separated by SDS–PAGE and a lane with equal intensity of overall Coomassie staining to the autophagosome separations was chosen for processing. Peptides were subjected to the C18 stage-tip method and resuspended in 5% formic acid, 5% acetonitrile before mass spectrometry analysis.

Peptides were separated on 100 μm × 25 cm C18 reversed phase (Maccel C18 3 μm 200 Å; The Nest Group) with a 90 min gradient of 6% to 27% acetonitrile in 0.125% formic acid. The twenty two most intense peaks from each full mass spectrometry scan acquired in the Orbitrap Velos Pro (Thermo) were selected for MS/MS (RAW files have been deposited at Peptide Atlas under accession number PASS00440). Sequest-based identification using a Human UNIPROT database followed by a target decoy-based linear discriminant analysis was used for peptide and protein identification as described[16]. Several experiments were processed in tandem using a protein sieve and protein assembler in-house processing tool, including PANC-1 chloroquine 4 h with PANC-1 Ex. 1 and PANC-1 Ex. 2, PANC-1 Ex. 3 and PANC-1 whole-cell lysate together. MCF7 Ex. 1, MCF7 Ex. 2, and MCF7 whole-cell lysate were processed together. However, all data sets were processed independently for calculation of Pearson correlation of $\log_2$(H:L) ratios between data sets for Fig. 1d and Extended Data Fig. 1j. Other parameters used for database searching include: 50 p.p.m. precursor mass tolerance; 1.0 Da product ion mass tolerance; tryptic or Lys-C digestion with up to three missed cleavages; and variable oxidation of Met (+15.994946). A protein level false-discovery rate of <1% was used as a threshold for protein identifications using the target decoy strategy. Quantification of each protein was determined using the peak heights for light and heavy forms for that protein. The criterion for protein quantification was a summed signal-to-noise ratio of >10. Quantification of protein level was by calculating the median value of the ratios of light to heavy. Processed mass spectrometry data including SILAC profiling information is available in Supplementary Tables 2–4, 6 and 7, with a guide to the experimental conditions for each experiment in Supplementary Table 1. Of note, contaminants including keratin, Lys-C or trypsin were removed from the data to avoid interference with data filtering as described later. In addition, peptides identified for MAP1LC3B (gene accession number 81631, NP_073729.1) are shared with a protein from an additional gene coding region with the gene symbol *MAP1LC3B2* (gene accession number 643246, NP_001078950.1). The protein assembler program reported MAP1LC3B2 as the protein identified. However there were no peptides to discriminate the two proteins. There is one amino acid difference between MAP1LC3B and MAP1LC3B2 (C113Y). But no peptides were identified within the region of C113Y (identified peptides are as follows: $_{31}$IPVIIER$_{37}$ and $_{52}$FLVPDHVNMSELIK$_{65}$). We therefore report the $\log_2$(H:L) ratios for MAP1LC3B in our figures and Extended Data Fig. 2 but leave the protein identifier as MAP1LC3B2 in our Supplementary Tables. The processed proteomics data reported in the paper are available in the Supplementary Information; RAW files have been deposited at Peptide Atlas under accession number PASS00440.

Bioinformatic analysis. Candidate autophagosomal proteins were identified by using a multi-step filtering process beginning with an enrichment cut-off including proteins with $\log_2$(heavy:light) ratios greater than 1.0 for MCF7 data sets, 1.5 for PANC-1 16 h data sets and 0.5 for PANC-1 and PA-TU-8988T 4 h chloroquine data sets (0.5 was used as a cut-off in the 4 h chloroquine data sets given the overall relatively lower level of maximum $\log_2$(heavy:light) ratios). To be included, proteins had to be represented by two or more peptides. At this point in the analysis, approximately 600 proteins remained per data set as potential candidates. To remove abundant proteins that may be non-specifically captured by bulk autophagy, candidates were filtered against the relative abundance of the proteome measured independently by LC–MS by directly comparing number of peptides identified per protein. Of note, a whole-cell lysate proteome data set was prepared from MCF7, PANC-1 and PA-TU-8988T cells as detailed earlier and was used for comparison within cell lines. In the data sets not processed together (PANC-1 4 h chloroquine and PANC-1 Ex. 1 versus PANC-1 whole-cell lysate) comparisons were done based on pairing of gene symbols that did not distinguish between isoforms within the same gene symbol. Finally, a two-sided Student's *t*-test was used as a measure of statistical confidence of the observed $\log_2$(heavy:light) ratio taking into account the standard

deviation of the $\log_2$(heavy:light) ratio and number of peptides measured per protein. A *P* value of less that 0.05 was used for inclusion as a candidate (except for the PANC-1 Ex. 1 data set that used Lys-C for digestion where a *P* value of less than 0.1 was used). For each data set, this typically left approximately 150 candidates. We subsequently determined the overlap between the three PANC-1 16 h chloroquine biological replicate data sets, identifying 86 proteins in common. The overlap between two MCF7 data sets consisted of 102 proteins. Finally, the overlap between the PANC-1 and MCF7 data sets was determined (33 proteins in common, 122 proteins specific to either PANC-1 or MCF7 data sets). This set of 155 proteins is referred to as class 1 candidate autophagosomal proteins. A subset of the top 50 candidate proteins, termed class 1A candidates, was developed based on presence in typically three or more independent experiments and those with known or potential links with autophagy. Finally, we also identified non-class-1 proteins with a $\log_2$(heavy:light) $>$ 2.0 in any two of the five independent PANC-1 (PANC-1 Ex. 1–3) or MCF7 (Ex. 1–2) profiling experiments (16 h in chloroquine), and will refer to this data set as class 2 autophagosome-enriched proteins (Supplementary Table 5).

For analysis of the semi-quantitative PA-TU-8988T GFP immunoisolation experiment, data was sorted by comparing $\log_2$ ratios of peptide numbers of proteins identified in autophagosomes purified from chloroquine versus wortmannin-treated cells and $\log_2$ ratios of peptide numbers of proteins identified in autophagosomes purified from chloroquine-treated cells versus peptide numbers from a whole-cell lysate sample (Supplementary Table 7). Zero value denominators were systematically replaced with a value of 0.5 to generate a $\log_2$ ratio. Candidates were qualified as enriched if both $\log_2$ ratios were greater than 0.5.

For preparation of Extended Data Fig. 5, the overlap between the MCF7 candidate proteins identified in this work and the stimulus independent autophagosomal candidates identified in Dengjel *et al.* was determined (table 1 from ref. 13). Only two proteins overlapped, namely SQSTM1 and GABARAPL2. To understand the lack of overlap between the data sets, we analysed the proteins from the Dengjel *et al.* list using the data from our MCF7 experiments. Of note, the Dengjel *et al.* experiments also used the MCF7 cell line. We curated the Dengjel *et al.* data sets to ensure the gene symbols published in their study matched the updated version of the database used in our proteomics analyses (Supplementary Table 8). For the 96 Dengjel *et al.* candidate proteins, we first determined the $\log_2$(heavy:light) ratio observed in both MCF7 Ex. 1 and 2 autophagosome enrichments (Supplementary Tables 4 and 8). A majority of the proteins had a $\log_2$(heavy:light) ratio greater than 1.0 (77 out of 96). However, of the 77 proteins with a $\log_2$(heavy:light) ratio greater than 1.0, only two proteins passed the whole proteome abundance filter used in our analysis to remove abundant proteins that may be non-specifically captured by bulk autophagy (Extended Data Fig. 5a and Supplementary Table 8). We subsequently included the three 16 h PANC-1 autophagosome proteomics data sets used to create the final class 1 and 2 lists to obtain as much coverage and comparison of the proteins identified in the Dengjel *et al.* data sets as possible. The overlap with the Dengjel *et al.* data set is not significantly improved even when we expanded the data to include our PANC-1 data sets (Extended Data Fig. 5b and Supplementary Table 8) now with an overlap of four proteins (SQSTM1, GABARAPL2, VPS35 and MAP1LC3B). As the Dengjel *et al.* authors derived their final autophagosomal candidate proteins from the overlap of lists from three different experiments, we also analysed the three experiments independently using the same approach as described earlier. Similarly, although a large proportion of the 'cluster A' proteins from each experiment were enriched by $\log_2$(heavy:light) ratio in our data sets, only a small number of these proteins passed the whole-proteome abundance filter and were found in our class 1 and 2 data set (Extended Data Fig. 5c–e). By relaxing the $\log_2$(heavy:light) ratio cut-off and redundancy stringency (reduced stringency column in Supplementary Table 8, which accompanies Extended Data Fig. 5b–e) to include non-class-1 and -2 proteins in the analysis, the additional overlap only ranges from 5–15% depending on the individual data set. Of note, the class 1–2 candidates that were also identified in the Dengjel *et al.* concanamycin A 'cluster A' proteins included NCOA4. The likely reason for the large number of proteins with a $\log_2$(heavy:light) ratio greater than 1.0 but that did not pass the whole-proteome abundance filter stems from the protein correlation profiling methodology used by Dengjel *et al.* Specifically, the authors relied on an autophagosomal enrichment profile that required identification and quantification of candidate proteins in all six iodixanol fractions for creation of an evaluable profile. This probably biased their identification towards proteins with a high abundance in the whole proteome that would be more likely to be identified in all six fractions. As described earlier, abundant proteins are more likely to be captured by non-selective bulk autophagy. Finally, by using the overlap of the three different experiments, they further biased their candidate list towards abundant proteins captured by autophagy under all three conditions tested, such as proteasome subunits and heat-shock proteins.

Electron microscopy. Pelleted autophagosome fractions (A1) were fixed in 2.5% glutaraldehyde, 1.25% paraformaldehyde and 0.03% picric acid in 0.1 sodium cacodylate buffer (pH 7.4) for 1 h at room temperature, washed in 0.1 M sodium cacodylate buffer (pH 7.4), and post-fixed for 30 min in 1% osmium tetroxide (OsO4)/1.5% potassium ferrocyanide ($KFeCN_6$). Autophagosome pellets were washed in water three times and incubated in 1% aqueous uranyl acetate for 30 min followed by two washes in water and subsequent dehydration in grades of alcohol (5 min each; 50%, 70%, 95%, two times 100%). Autophagosome pellets were infiltrated overnight in a 1:1 mixture of propyleneoxide and TAAB Epon (Canemco & Marivac). The samples were subsequently embedded in TAAB Epon and polymerized at 60 °C for 24–48 h. Ultrathin sections (60–80 nm) were cut on a Reichert Ultracut-S microtome, picked up on to copper grids stained with lead citrate and examined using a Tecnai G2 Spirit BioTWIN transmission electron microscope and images were recorded with an AMT 2k CCD camera.

Interaction proteomics. Interaction proteomics was performed essentially as described previously[22], but with small modifications. Briefly, 293T, PANC-1 or PA-TU-8988T cells were transduced with a lentiviral vector expressing NCOA4–HA–Flag (NP_001138735.1) or FTH1–HA–Flag (293T only, NP_002023.2) and stable cell lines were selected in puromycin. Cells from 4×15 cm dishes at 80% confluence were harvested and lysed in 3 ml of 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5% Nonidet P40, 2 mM dithiothreitol (DTT) and protease inhibitors. Cleared lysates were filtered through 0.45 µm spin filters (Millipore Ultrafree-CL) and immunoprecipitated with 30 µl anti-HA resin (Sigma). Complexes were washed four times with lysis buffer, exchanged into PBS for a further three washes, eluted with HA peptide and precipitated with 10% TCA. TCA-precipitated proteins were trypsinized, purified with Empore C18 extraction media (3 M), and analysed via LC–MS/MS with a LTQ-Velos linear ion trap mass spectrometer (Thermo) with an 18 cm^3 125 µm (ID) C18 column and a 50 min 8–26% acetonitrile gradient. All AP–MS experiments in 293T cells were performed in biological duplicate and for each biological experiment, complexes were analysed twice by LC–MS/MS to generate technical duplicates. AP–MS experiments in PANC-1 and PA-TU-8988T cells were performed on a single AP but with technical duplicates. Spectra were searched with Sequest against a target-decoy human tryptic UNIPROT-based peptide database, and these results were loaded into the Comparative Proteomics Analysis Software Suite (CompPASS) to identify HCIPs[22]. Here a statistics table, derived from analogous AP–MS data for 172 unrelated proteins in 293T cells was used to determine WDN scores as well as *Z* scores based on spectral counts. The PANC-1 data were analysed using a PANC-1-specific statistics table with 12 unrelated AP–MS PANC-1 experiments and the PA-TU-8988T AP–MS experiment was analysed using the 293T statistics table given that no statistics table was available for the PA-TU-8988T cell line. The *D* score measures the reproducibility, abundance and frequency of individual proteins detected in each individual analysis. To identify NCOA4-associated proteins, we filtered proteins at a 2% false discovery rate for those with a WDN score $\geq$ 1.0 and average assembled peptide spectral matches (APSMs) $\geq$ 2 in both biological duplicates. Data presented in Fig. 2e are derived from the 293T AP–MS experiments and the figure was made using Cytoscape[29]. STRING database data are represented as previously[22]. FTH1 and FTL interactors were confirmed in both PANC-1 and PA-TU-8988T AP–MS experiments and HERC2 in the PA-TU-8988T AP–MS. NEURL4 was included in the interaction network given identification in the 293T AP–MS experiment and previously published data revealing a HERC2–NEURL4 interaction[23].

Immunological methods and microscopy. To assess enrichment of autophagosome-associated proteins in autophagosome purifications, the gradient load and autophagosome fractions were extracted in 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% Nonidet P40, 0.1% SDS, 2 mM EDTA, 2 mM DTT, protease inhibitors and subjected to immunoblotting with the indicated antibodies. To assess autophagosome integrity, purified autophagosome fractions (A1) were incubated in buffer A for 1 h at 37 °C plus or minus 0.5% Triton X-100, centrifuged at 16,000*g* for 10 min, and the supernatant and pellet were subjected to immunoblotting with the indicated antibodies. To validate interactions between NCOA4 and candidate interacting proteins, 293T cells stably expressing NCOA4–HA–Flag were harvested at 80% confluency. Extracts (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5% Nonidet P40, 2 mM DTT and protease inhibitors) from cells were subjected to immunoprecipitation with anti-Flag resin (M2 agarose; Sigma), and washed complexes subjected to immunoblotting with the indicated antibodies. Likewise, 293T cells stably expressing FTH1–HA–Flag with transient expression of MYC–NCOA4 were harvested at 80% confluency, lysed as described earlier, subjected to immunoprecipitation as described earlier, and immunoblotted with the indicated antibodies.

To examine localization of high-priority candidate proteins, full-length clones of the indicated candidate proteins were stably expressed as HA–Flag fusions in PA-TU-8988T and U2OS cells. Cells were plated on glass coverslips, treated with vehicle or chloroquine (10 µM, 4 h), and fixed with 4% paraformaldehyde before immunofluorescence using anti-HA to detect candidate proteins and anti-LC3B to detect autophagosomes. All images were collected with a Yokogawa CSU-X1 spinning disk confocal with Borealis modification on a Nikon Ti-E inverted microscope equipped with ×100 Plan Apo NA 1.4 objective lens. HA-candidate protein

fluorescence was excited with the 488 nm line (selected with an AOTF) from Spectral Applied Precision LMM-7 solid state laser launch. Emission was collected with a quadruple band pass polychroic mirror (Semrock) and a Chroma ET525/50 m emissions filter. LC3B fluorescence was excited with the 561 nm line from the LMM-7 launch, and emission collected with the Semrock polychroic and a Chroma ET620/60 m emission filter. For triple co-localization (Fig. 3c), ferritin fluorescence was excited with the 642 nm line from the LMM-7 launch, and emission collected with the Semrock polychroic and a Chroma ET700/75 m emission filter. Images were acquired with a Hamamatsu ORCA-AG cooled CCD camera controlled with MetaMorph 7 software. Z-series optical sections were collected with a step size of 0.2 μm, using the internal Nikon Ti-E focus motor. Co-localization was determined based on examination of single z slices using MetaMorph 7 software (results are tabulated in Extended Data Fig. 2a).

To examine NCOA4 protein levels while blocking autophagic flux, PA-TU-8988T cells were treated with vehicle (PBS or DMSO), chloroquine (10 μM, 8 h) or BAF (50 nM, 8 h), extracted in 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% Nonidet P40, 0.1% SDS, 2 mM EDTA, 2 mM DTT, protease inhibitors and subjected to immunoblotting with the indicated antibodies.

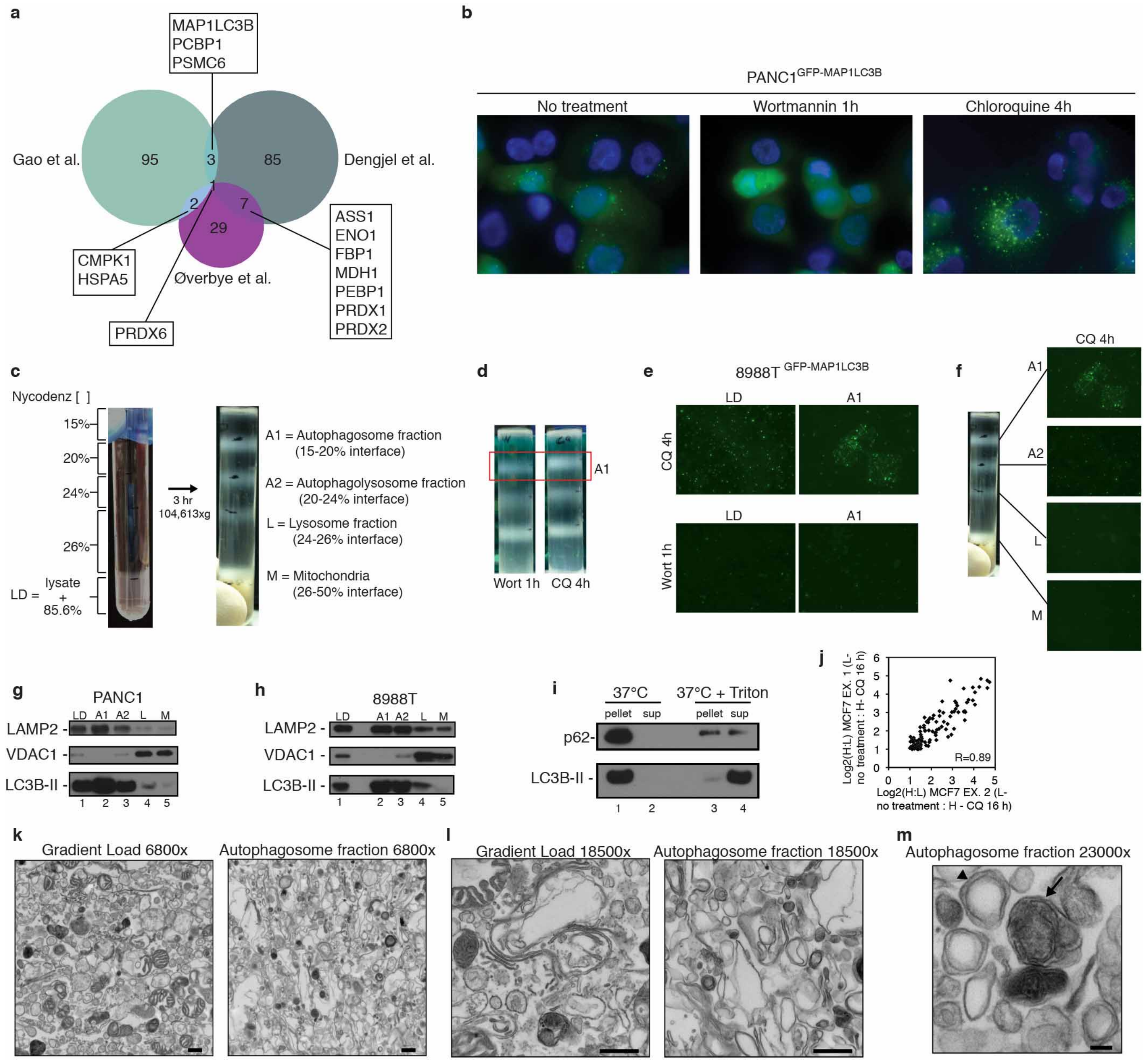
To examine co-localization of NCOA4 with LC3B and ferritin, an amino-terminal GFP fusion of NCOA4 was used. U2OS and PA-TU-8988T cells stably expressing GFP–NCOA4 were treated with vehicle (PBS or DMSO), chloroquine (10 μM, 4 h), BAF (50 nM, 4 h) or FAC (36 μM or 180 μM, 24 h) and fixed with 4% paraformaldehyde before immunofluorescence using anti-LC3B to label autophagosomes or anti-ferritin in FAC-loaded cells. Images were collected as described earlier with 6–10 z-series optical sections collected with a step size of 0.2 μm, using the internal Nikon Ti-E focus motor. Z-series are displayed as maximum z-projections.

GST pull-down. The six human paralogue ATG8 proteins were produced as N-terminal GST fusions (GST–ATG8) and transformed in *Escherichia coli* BL21 (DE3) cells. Protein expression was induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG) for 3 h at 37 °C. Cells were collected by centrifugation, re-suspended in 25 mM Tris-HCl (pH 7.4), 10% (w/v) sucrose and lysed with a single freeze–thaw cycle followed by sonication. After centrifugation, lysates were incubated with 10 μl of a 50% (v/v) slurry of glutathione Sepharose 4B beads for 30 min at 4 °C. Beads were washed thoroughly with 150 mM NaCl, 20 mM Tris (pH 7.4), 2 mM DTT, 0.5% (w/v) Nonidet P-40, leaving 50–100 μg of GST-fusion protein bound to the beads. Beads were then mixed with 300 μg of lysate from a NCOA4–HA–Flag stably expressing 293T cell line. The assay mix was incubated for 30 min at 4 °C, and beads were washed four times with 1 ml wash buffer. Proteins were eluted with SDS sample buffer and analysed by 4–20% gradient SDS–PAGE and immunoblotting with HA antibody. Pounceau stain was used to visualize GST–ATG8 bands.

Chelation assays. Chelation assays were performed similarly to those previously described[5] with slight modifications. Briefly, U2OS and PA-TU-8988T cells stably expressing shRNAs as described earlier (shGFP, shNCOA4-1, shNCOA4-2, selected with puromycin) were plated (2×10^5 cells per well) and cultured for 24 h in FAC (36 μM or 180 μM). Cells were washed three times with PBS and subjected to iron chelation with the described chelators minus or plus lysosomal protease inhibitors (E-64d and pepstatin A) or the proteasomal inhibitor, bortezomib. Cells were harvested at the indicated time points, washed with PBS, and lysed in 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% Nonidet P40, 0.1% SDS, 2 mM EDTA, 2 mM DTT, protease inhibitors and analysed by 4–20% gradient SDS–PAGE followed by immunoblotting with the indicated antibodies. Experiments were performed at a minimum in biological triplicate. Quantification of western blots was performed using ImageJ[30]. Cells were treated similarly for immunofluorescence experiments (including IMR-90 and HPDE cells). Cells plated on glass coverslips were treated as described above and fixed with 4% paraformaldehyde before immunofluorescence using anti-ferritin and anti-LAMP2. Data were collected as described earlier. For quantification of ferritin localization, at least five ×40 fields were collected for biological replicate experiments. The fraction of punctate ferritin was quantified by measuring the percentage of punctate ferritin immunofluorescence divided by total cellular ferritin immunofluorescence using MetaMorph software. At least five fields were quantified from biological duplicate experiments. For Fig. 3d, the number of PA-TU-8988T cells quantified expressing shGFP was 155, for PA-TU-8988T cells expressing shNCOA4-1 it was 113, and for shNCOA4-2 it was 138. For Extended Data Fig. 8c, the number of U2OS cells quantified is as follows: shGFP, 133 cells; shNCOA4-1, 103 cells; and shNCOA4-2, 79 cells. For Extended Data Fig. 8d, the number of IMR-90 cells quantified is as follows: shGFP, 29 cells; shNCOA4-1, 26 cells; shNCOA4-2, 31 cells. For Extended Data Fig. 9a, the number of PA-TU-8988T cells quantified is as follows: PA-TU-8988T-control MSCV-shGFP, 100 cells; PA-TU-8988T-control-MSCV-shNCOA4-1, 125 cells; PA-TU-8988T-control-MSCV-shNCOA4-2, 132 cells; PA-TU-8988T-mouse-NCOA4-shGFP, 151 cells; PA-TU-8988T-mouse-NCOA4-shNCOA4-1, 153 cells; and PA-TU-8988T-mouse-NCOA4-shNCOA4-2, 172 cells. Exposure times were held constant between experiments. Rescue experiments were performed using retroviral (MSCV blasticidin)-based expression of the murine homologue of NCOA4 (NP_001029160.1) with an N-terminal HA tag.

H_2O_2 assay. Experiments were performed as described previously[26] with slight modifications. PA-TU-8988T cells stably expressing shGFP or shNCOA4-1 were plated on a 96-well plate and treated with increasing concentrations of H_2O_2 as indicated for 30 min followed by 72 h of culturing in complete media. Relative cell viability was measured using CellTiter-Glo. Experiments were performed at least three times, each time in technical triplicate.

29. Lopes, C. T. *et al.* Cytoscape Web: an interactive web-based network browser. *Bioinformatics* **26,** 2347–2348 (2010).
30. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9,** 671–675 (2012).



Extended Data Figure 1 | Autophagosome enrichment protocol validation. **a**, Overlap of proteins identified in three prior autophagosome proteomics studies[11–13] as evaluated by area-proportional Venn diagram. Proteins overlapping between data sets are noted. **b**, PANC-1 cells stably expressing GFP–LC3B have a high level of basal autophagy (left). One hour of wortmannin (Wort; 200 nM) treatment blocks autophagosome formation (middle) and 4 h of chloroquine treatment (CQ; 25 μM) causes accumulation of autophagosomes (right). **c**, Lysed PA-TU-8988T cells mixed with Nycodenz and placed at the bottom of a discontinuous density gradient with Nycodenz layers at the indicated concentrations (left). After 3 h centrifugation at the indicated speed, four bands appear at the indicated interfaces with enrichment of the indicated organelles in each interface (right). **d**, PA-TU-8988T cells treated with either wortmannin (1 h, 200 nM) or chloroquine (4 h, 25 μM) subjected to gradient centrifugation. A decreased amount of material is recovered from the A1 (autophagosome) interface due to the effect of wortmannin on autophagosome formation. **e**, Fluorescence microscopy of gradient load (LD) and autophagosome fraction (A1) from PA-TU-8988T cells stably expressing GFP–LC3B after either chloroquine or wortmannin treatment. **f**, Fluorescence microscopy of indicated fractions from density gradient of PA-TU-8988T cells stably expressing GFP–LC3B treated with chloroquine (A1 fraction image is also presented in **e** and gradient picture is also presented in **c**). **g**, PANC-1 autophagosome fractions analysed by immunoblotting using antibodies to LAMP2, VDAC1 and LC3B. LD is gradient load, A1 is autophagosome fraction from 15–20% nycodenz interface, A2 is the autophagolysosome fraction from the 20–24% nycodenz interface, L is the lysosome fraction from the 24–26% nycodenz interface, M is the mitochondrial fraction from the 26–50% nycodenz interface. **h**, PA-TU-8988T autophagosome fractions analysed as in **g**. **i**, PA-TU-8988T autophagosomes (A1 fraction) were incubated at 37 °C for 1 h ± Triton X-100 and centrifuged at high speed. The resulting pellet was resuspended in equal volume to supernatant and assayed by immunoblotting with antibodies to SQSTM1 and LC3B. **j**, Pearson correlation plot for overlapping candidates from MCF7 experiments (102 proteins, comparing Ex. 1 with Ex. 2). **k**, Electron micrographs of PA-TU-8988T gradient load (LD, left) and PA-TU-8988T autophagosome fraction (A1, right) at ×6,800 magnification. Scale bar, 500 nm. **l**, PA-TU-8988T gradient load (LD, left) and PA-TU-8988T autophagosome fraction (A1, right) at ×18,500 magnification. Scale bar, 500 nm. **m**, PA-TU-8988T autophagosome fraction (A1) at ×23,000 magnification. Scale bar, 100 nm. Arrowhead indicates double-membrane autophagosome, arrow indicates fused autophagolysosome.

a

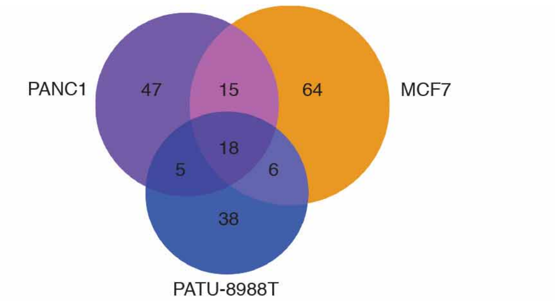
Gene Symbols	ID*	IF	PANC1 Ex. 3 No. Peptides	PANC1 Ex. 3 Log2(H:L)
SQSTM1	R		78	5.01
GABARAPL2	A		9	4.97
TAX1BP1	C	Y	50	4.73
CALCOCO2	R		43	4.56
STX17	M		3	4.50
MAP1LC3B	A		3	4.38
NCOA4	C	Y	11	4.37
KEAP1	M		14	4.29
ITM2B	U		30	4.20
NBR1	R		12	4.17
APP	S		93	3.99
APLP2	S		182	3.99
ITM2C	U		26	3.93
CALCOCO1	C	Y	8	3.93
SDC4	C	Y	24	3.84
TMEM59	R		32	3.80
CD320	U		7	3.80
IFNGR1	U		10	3.74
CLU	C	N	20	3.65
FAT1	U		138	3.57
SNX17	U		13	3.56
NDFIP1	C	Y	19	3.52
OPTN	R		6	3.37
NOTCH2	U		32	3.30
LAPTM4A	M		17	3.24
RNF130	U		2	3.17
UGCG	U		17	3.14
PRKACA	M		4	2.99
RNF149	C	Y	11	2.98
ITCH	M		12	2.88
GPR56	U		5	2.70
TMF1	C	Y	11	2.65
OSMR	U		13	2.63
FYCO1	M		13	2.61
NRP1	S		24	2.57
ANTXR1	U		13	2.51
IL6ST	C	N	35	2.46
JAK1	C	N	28	2.42
TNFRSF10B	U		12	2.25
VPS26A	H		20	2.16
LGALS8	M		15	1.93
SLC38A2	U		44	1.93
STX6	H		19	1.92
MFGE8	U		69	1.90
MGRN1	C	N	9	1.82
RAB7A	H		84	1.33
SPTLC1	H		13	1.10
NAPA	H		37	1.01
SLC7A1	H		23	1.00
ATP6V1H	H		17	0.50

MCF7 | 8988T SILAC | 8988T LC3-IP

*ID (Identification):
R = cargo adaptor
A = ATG8
C = candidate tested for colocalization
U = candidate not tested for colocalization
M = autopahgy machinery
S = selective autophagy cargo
H = found as HCIP in autophagy interaction network

Enriched
Not enriched
Not identified

b

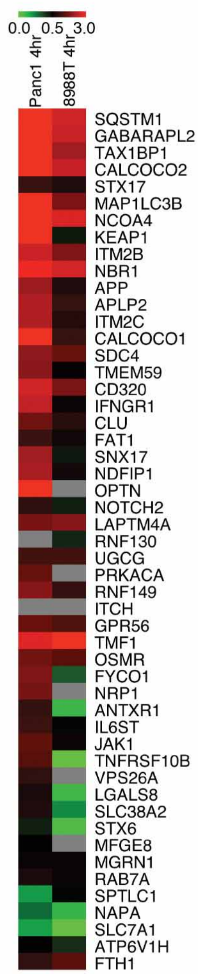


c

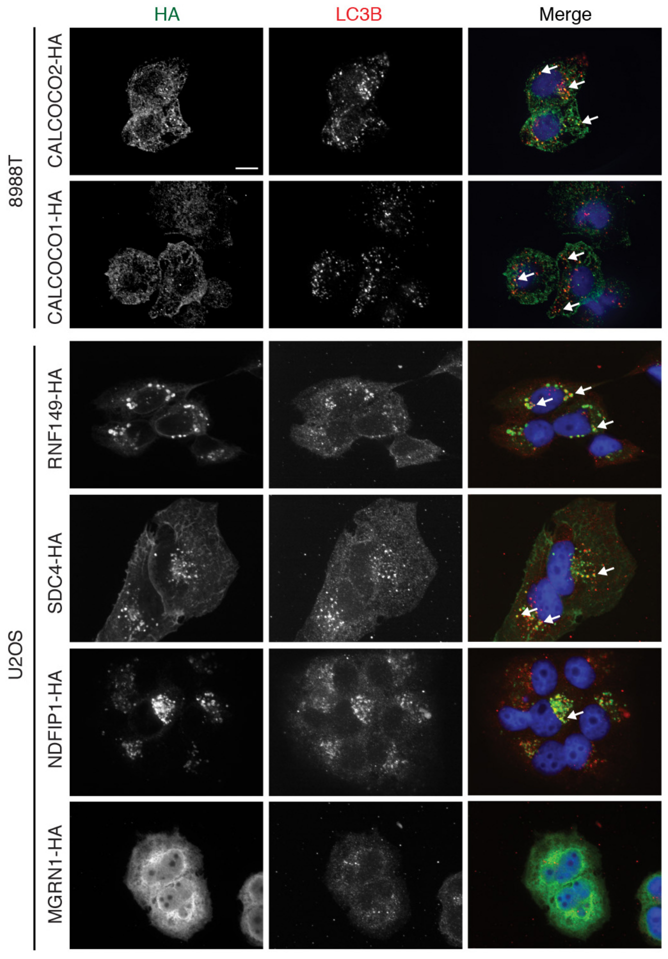
Gene Symbol	8988T Peptide #	8988T log2(H:L)
SQSTM1	113	2.57
GABARAPL2	19	2.50
TAX1BP1	14	2.09
CALCOCO2	18	2.50
STX17	6	0.80
MAP1LC3B	8	1.74
NCOA4	12	2.66
KEAP1	6	0.46
ITM2B	14	1.78
NBR1	16	2.60
APP	54	0.79
APLP2	142	1.09
ITM2C	21	0.84
CALCOCO1	6	1.08
SDC4	21	1.53
TMEM59	22	0.55
CD320	12	1.73
IFNGR1	5	0.60
CLU	6	0.92
FAT1	135	0.65
SNX17	12	0.46
NDFIP1	13	0.67
OPTN	0	0
NOTCH2	12	0.45
LAPTM4A	10	1.84
RNF130	4	0.43
UGCG	19	1.22
PRKACA	0	0
RNF149	8	1.06
ITCH	0	0
GPR56	1	1.33
TMF1	1	3.63
OSMR	23	1.42
FYCO1	18	0.33
NRP1	0	0
ANTXR1	7	0.11
IL6ST	24	0.49
JAK1	22	0.59
TNFRSF10B	8	-0.08
VPS26A	0	0
LGALS8	21	0.11
SLC38A2	62	0.23
STX6	53	0.07
MFGE8	0	0
MGRN1	9	0.58
RAB7A	103	0.60
SPTLC1	23	0.49
NAPA	42	0.12
SLC7A1	55	0.01
ATP6V1H	11	0.42
FTH1	11	1.42

Enriched
Not enriched
Not identified

d

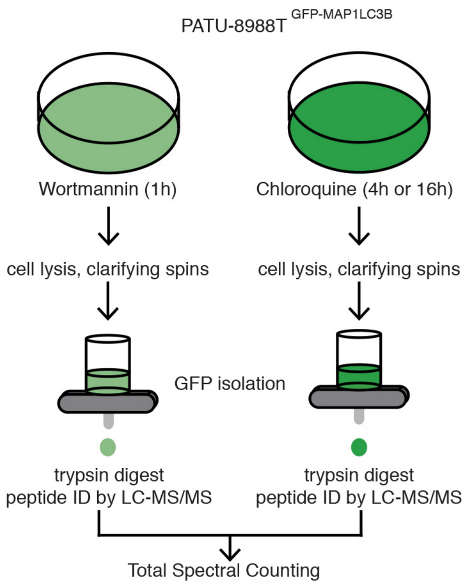


Extended Data Figure 2 | Autophagosome proteomics candidate list and validation in PA-TU-8988T cells. **a**, Autophagosome proteomics class 1A candidate list. **b**, Overlap between class 1 candidates (MCF7 and PANC-1) and candidates from 4 h chloroquine SILAC PA-TU-8988T gradient autophagosome purification experiment. **c**, Data from PA-TU-8988T SILAC gradient autophagosome for class 1A candidates (and FTH1; see also Supplementary Table 6). **d**, Heat map of class 1A candidates (and FTH1) comparing PANC-1 4 h chloroquine and PA-TU-8988T 4 h chloroquine experiments.



Extended Data Figure 3 | Candidates from autophagosome proteomics co-localize in autophagosomes. Representative confocal images of PA-TU-8988T and U2OS cells expressing HA-tagged candidates after chloroquine treatment. Co-localization of HA-tagged candidates (green) with endogenous LC3B (red). Representative co-localization is marked by white arrows. Scale bar, 10 μm. MGRN1 is included as an example of one of the candidates that did not show co-localization.

a

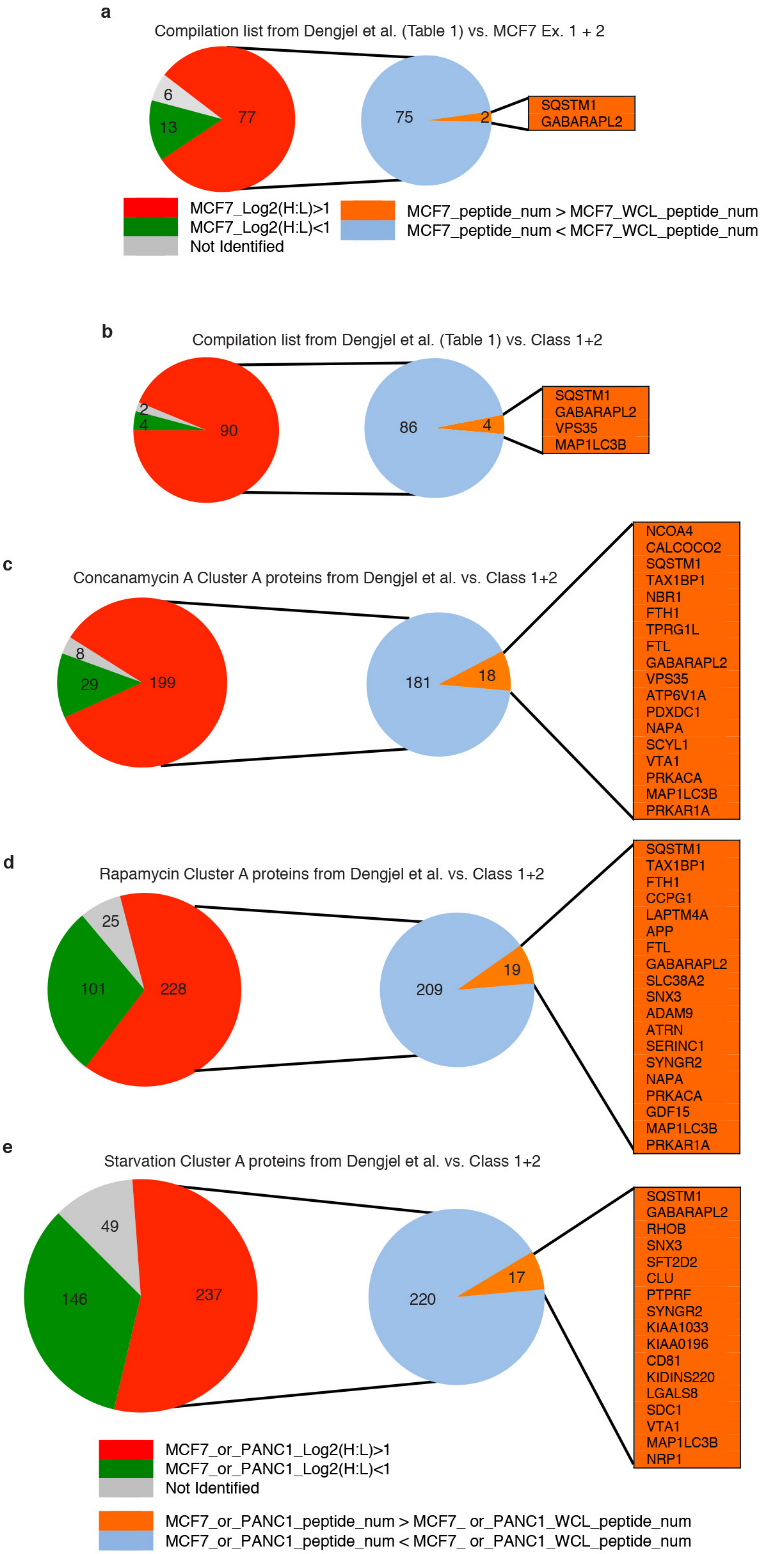


b

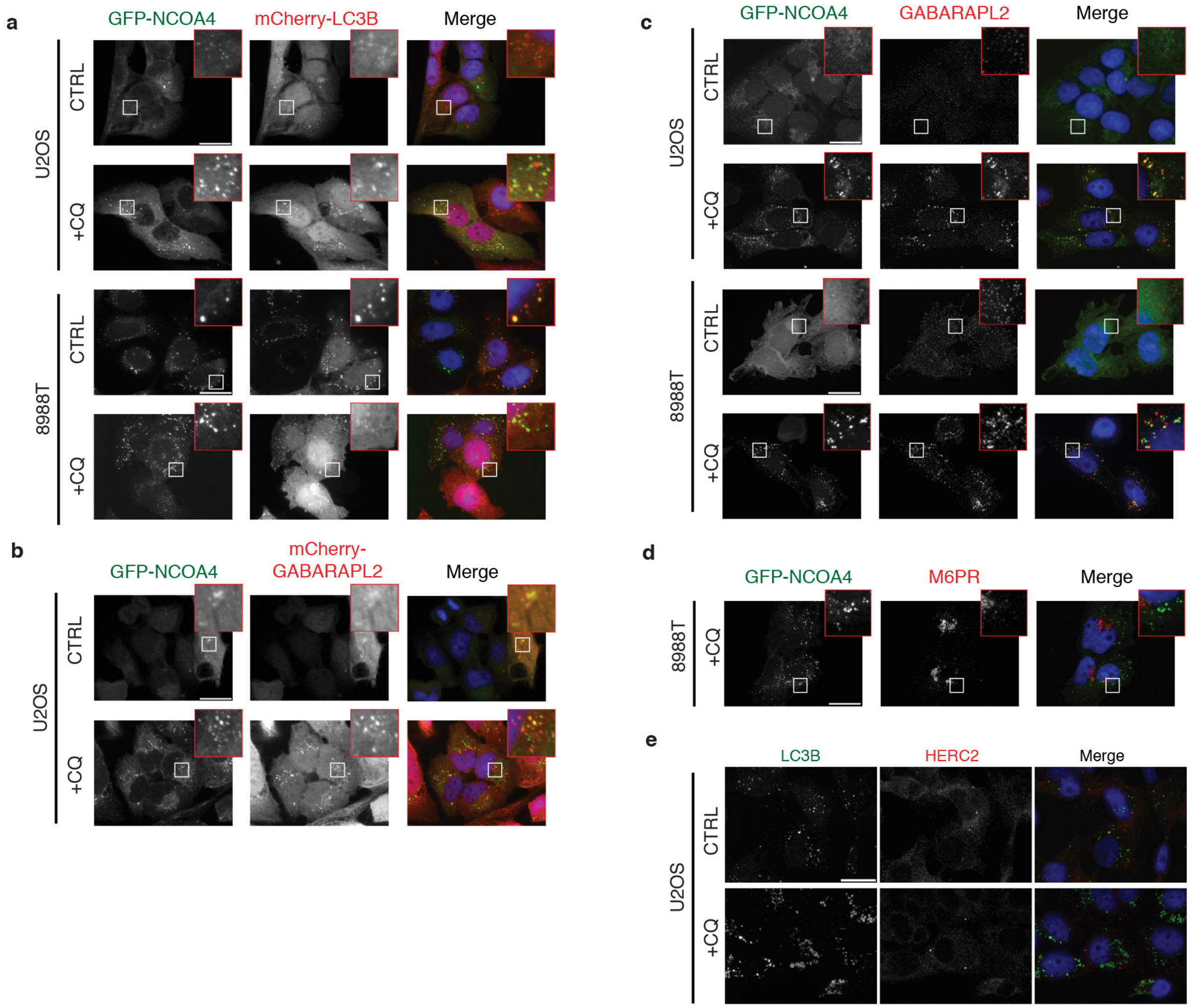
Gene Symbols	Log2(CQ_peptide#: WORT_peptide#)	Log2(CQ_peptide#: WCL_peptide#)
SQSTM1	3.79	3.60
GABARAPL2	3.17	3.17
TAX1BP1	4.00	4.00
CALCOCO2	0.00	0.00
STX17	0.00	0.00
MAP1LC3B	0.00	0.00
NCOA4	4.09	5.09
KEAP1	3.32	3.32
ITM2B	0.70	4.70
NBR1	3.58	3.58
APP	0.75	6.00
APLP2	0.18	7.41
ITM2C	0.85	3.17
CALCOCO1	3.00	2.00
SDC4	0.58	3.58
TMEM59	2.00	5.00
CD320	0.00	0.00
IFNGR1	0.58	3.58
CLU	0.58	2.58
FAT1	-0.39	6.67
SNX17	3.00	3.00
NDFIP1	1.58	2.58
OPTN	3.00	3.00
NOTCH2	2.17	4.17
LAPTM4A	2.00	4.00
RNF130	1.00	2.00
UGCG	0.87	4.46
PRKACA	2.58	2.58
RNF149	2.32	3.32
ITCH	3.32	2.32
GPR56	-2.00	0.00
TMF1	2.58	2.58
OSMR	0.36	5.17
NRP1	0.00	0.00
ANTXR1	0.19	4.00
IL6ST	0.76	4.46
JAK1	1.00	3.58
TNFRSF10B	-0.42	2.58
VPS26A	2.58	-1.58
LGALS8	-1.00	-1.00
SLC38A2	1.23	5.32
STX6	0.53	3.70
MFGE8	0.00	0.00
MGRN1	-0.58	2.00
FYCO1	0.00	0.00
RAB7A	0.77	1.83
SPTLC1	-1.22	0.00
NAPA	0.78	0.26
SLC7A1	0.22	2.22
ATP6V1H	0.00	-1.00

Enriched
Not enriched
Not identified

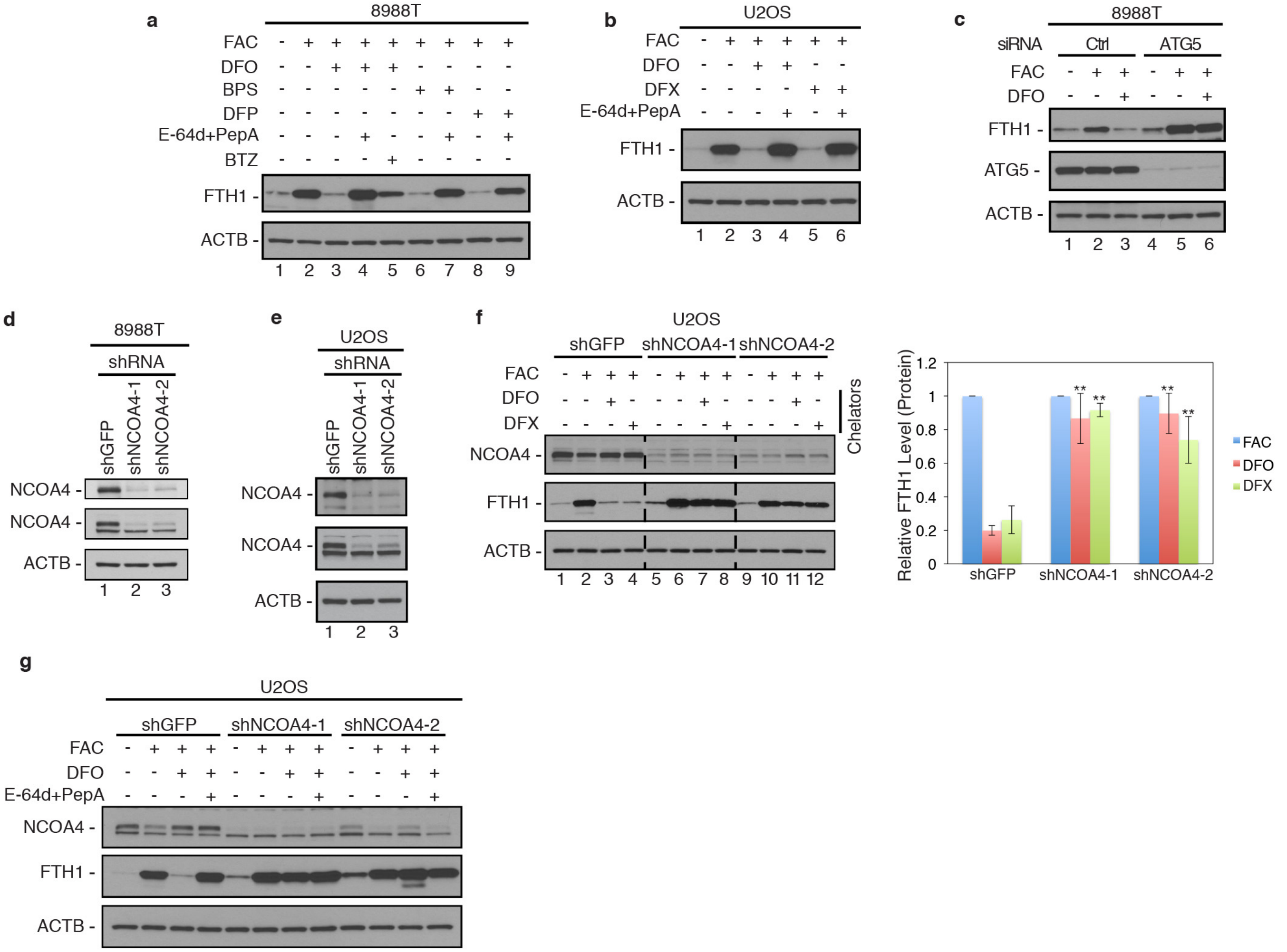
Extended Data Figure 4 | Immunopurification-based autophagosome proteomics. **a**, Schematic for GFP immunoisolation of GFP–LC3B-labelled autophagosomes from PA-TU-8988T cells. **b**, Data from GFP immunoisolation for class 1A candidates (see also Supplementary Table 7).



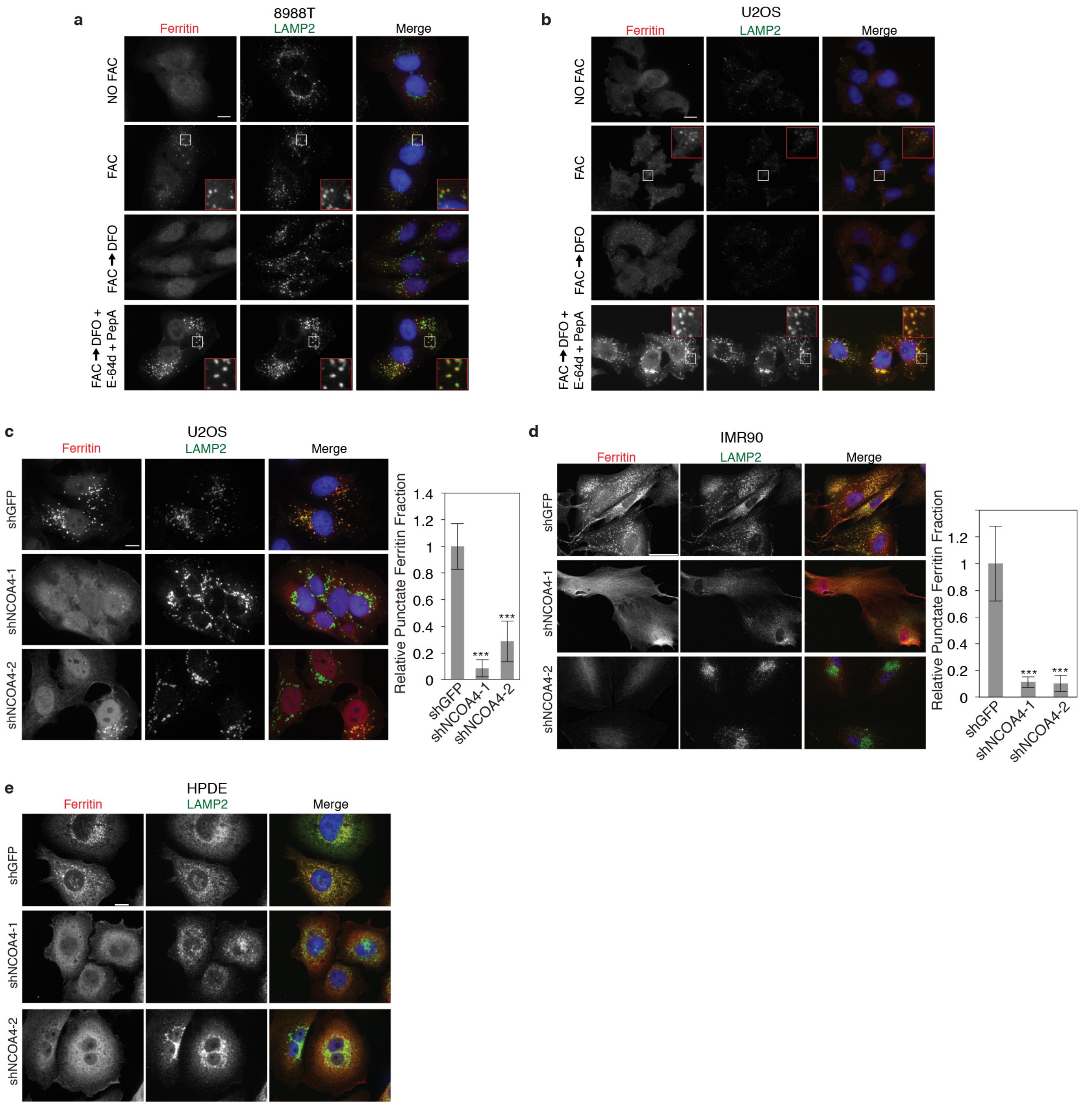
Extended Data Figure 5 | Comparative analysis of class 1 and 2 proteins with Dengjel *et al.* autophagosome proteomics. **a**, Analysis and comparison of Dengjel *et al.*[13] data and candidate list with data derived from MCF7 autophagosome proteomics experiments as detailed in Methods (see also Supplementary Table 8). Orange shading of gene symbols denotes proteins identified as MCF7 candidates. **b**–**e**, Analysis and comparison of Dengjel *et al.* data and candidate list with data derived from both MCF7 and PANC-1 autophagosome proteomics experiments (class 1 and 2 proteins) as detailed in Methods. Orange shading of gene symbols denotes proteins identified as class 1 or 2 candidates (see also Supplementary Table 8).



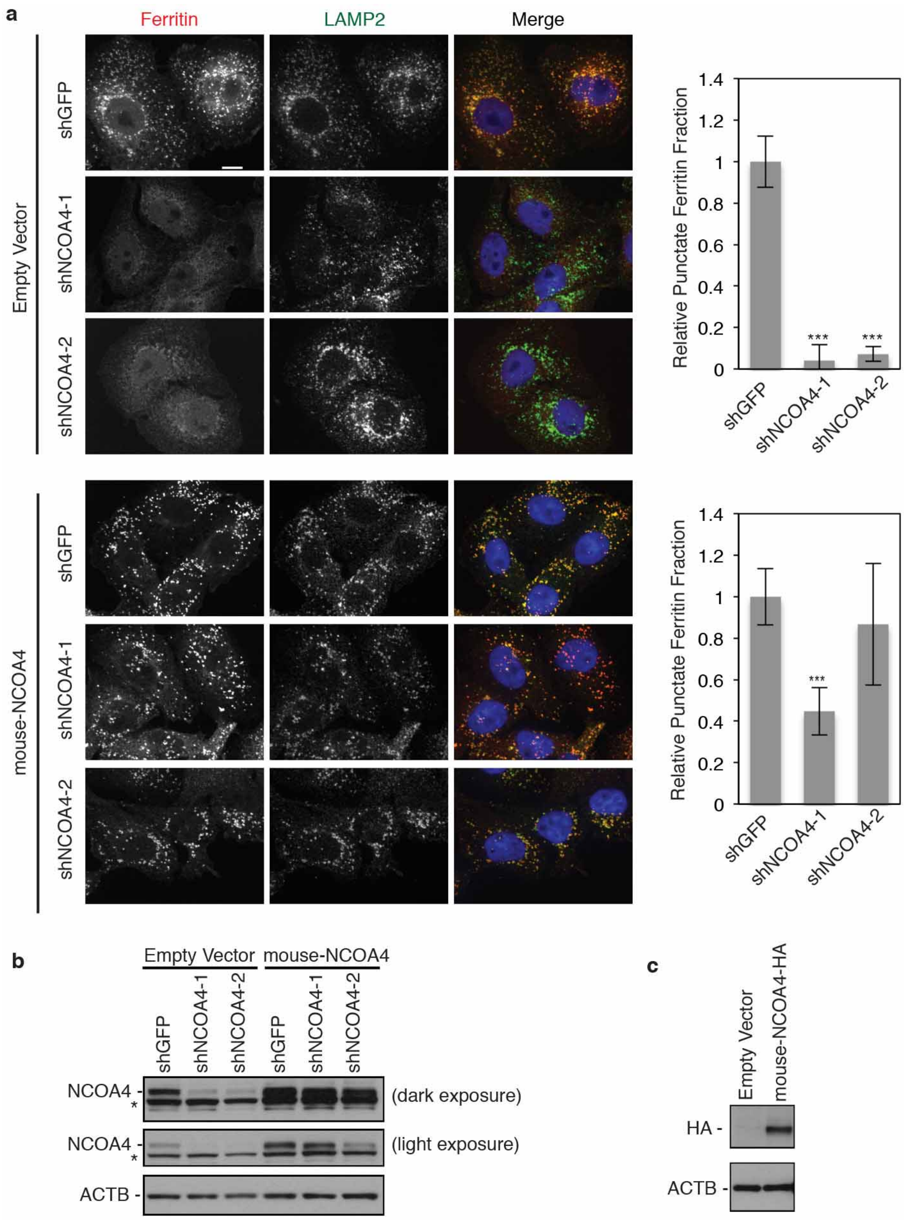
Extended Data Figure 6 | NCOA4 co-localizes with LC3B and GABARAPL2 in autophagosomes. **a**, GFP–NCOA4 (green) co-localizes with mCherry–LC3B (red) in chloroquine-treated cells. Scale bar, 20 µm. **b**, GFP–NCOA4 (green) co-localizes with mCherry–GABARAPL2 (red) in chloroquine-treated cells. Scale bar, 20 µm. **c**, GFP–NCOA4 (green) co-localizes with endogenous GABARAPL2 (red) in chloroquine-treated cells. Scale bar, 20 µm. **d**, GFP–NCOA4 (green) does not co-localize with endogenous mannose 6-phosphate receptor (M6PR) (red) in chloroquine-treated cells. Scale bar, 20 µm. **e**, HERC2 does not co-localize in autophagosomes. Immunostaining of U2OS cells subjected to chloroquine treatment, endogenous LCB3 (green) and endogenous HERC2 (red). Scale bar, 20 µm.



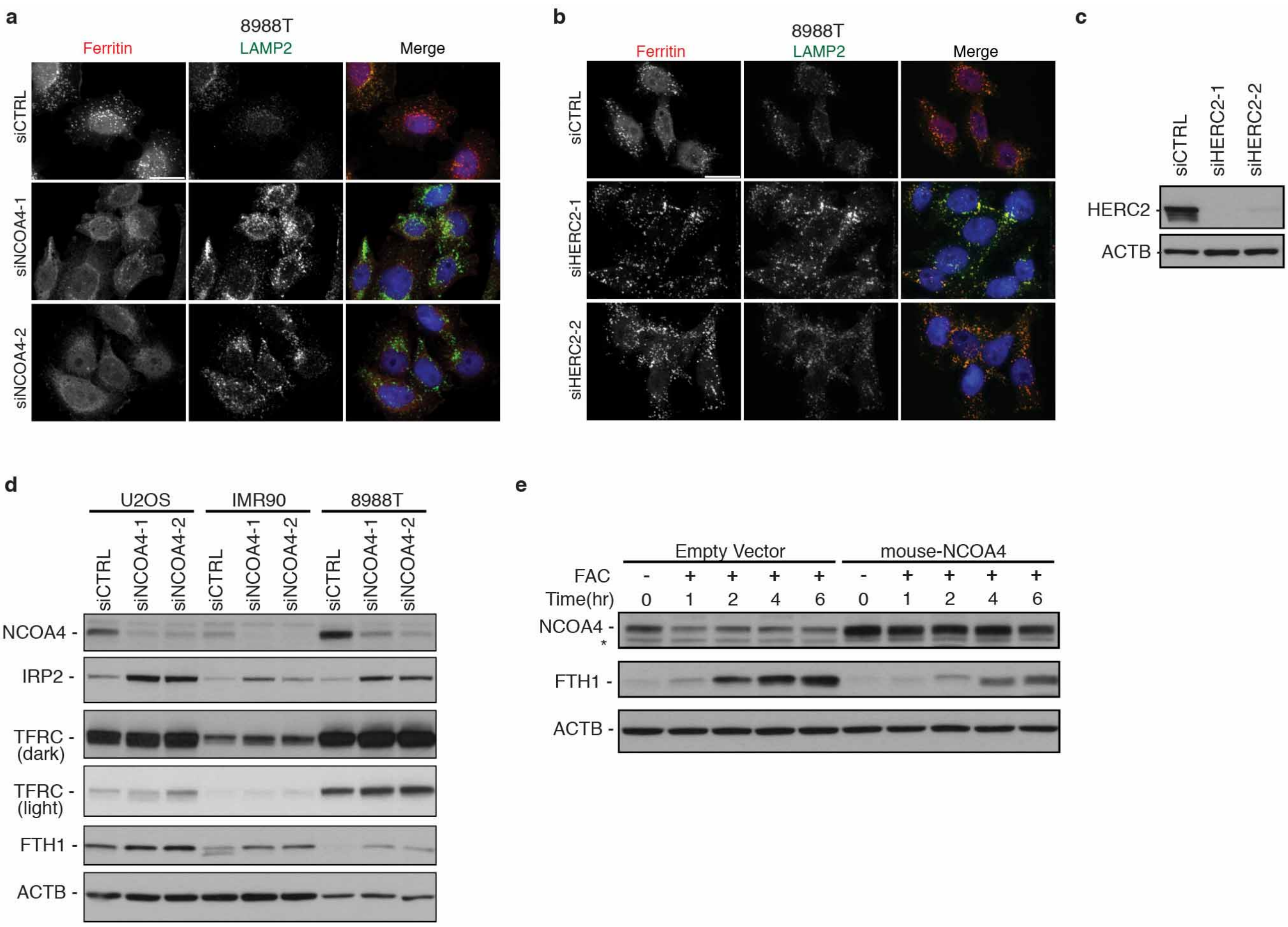
Extended Data Figure 7 | Ferritin undergoes primarily lysosomal mediated degradation. **a**, PA-TU-8988T cells were cultured with FAC for 24 h, washed, followed by chelation with three chelators (DFO, BPS, DFP) plus or minus lysosomal protease inhibitors (E-64d and pepstatin A) or proteasomal inhibitor, bortezomib (lane 5) for 8 h. Lysates were immunoblotted using antibodies to ACTB and FTH1. **b**, U2OS cells were cultured with FAC for 24 h, washed, followed by chelation with two chelators (DFO, DFX) plus or minus lysosomal protease inhibitors (E-64d and pepstatin A) and analysed as in **a**. **c**, PA-TU-cells transfected with luciferase control siRNA or validated siRNA to ATG5 were cultured with FAC, washed and subjected to DFO chelation for 9 h. Lysates were immunoblotted using antibodies to FTH1, ATG5 and ACTB. **d**, RNAi-mediated knockdown of NCOA4 in PA-TU-8988T cells. PA-TU-8988T cells stably expressing a control shRNA (shGFP) and two independent shRNAs to NCOA4 (shNCOA4-1 and shNCOA4-2) were lysed and analysed by immunoblotting with two different antibodies to NCOA4 and ACTB as a loading control. Middle panel shows immunoblot probed with NCOA4 antibody from Bethyl Laboratories (#A302-272A). A non-specific band migrates just below the NCOA4-specific band. Top panel shows immunoblot probed with NCOA4 antibody from Sigma (SAB1404569). **e**, U2OS cells stably expressing shGFP, shNCOA4-1 or shNCOA4-2 were analysed by immunoblotting as in **d**. **f**, NCOA4 depletion rescues ferritin degradation upon 9 h iron chelation in U2OS cells. Relative FTH1 levels ($n = 3$) for each chelator are quantified. Bars and error bars represent mean values and s.d., respectively. $^{**}P < 0.01$ and $^{*}P < 0.02$ comparing FTH1 levels between different cell lines to shGFP control (one-sided t-test). **g**, U2OS cells stably expressing shGFP, shNCOA4-1 or shNCOA4-2 were cultured with FAC for 24 h, washed, and subjected to DFO chelation plus or minus lysosomal protease inhibitors. Lysates were immunoblotted using antibodies to NCOA4, FTH1 and ACTB.



Extended Data Figure 8 | NCOA4 mediates autophagic delivery of ferritin to lysosomes. **a**, PA-TU-8988T cells were cultured in the presence or absence of FAC for 24 h, washed, and subjected to DFO chelation plus or minus lysosomal protease inhibitors (E-64d and pepstatin A). Cells were fixed and immunostained using antibodies to ferritin (red) and LAMP2 (green). Higher magnification views of the boxed areas are shown in the insets. Scale bar, 10 μm. **b**, U2OS cells treated and analysed as in **a**. **c**, Immunostaining of U2OS cells expressing a control shRNA (shGFP) and two independent shRNAs to NCOA4 (shNCOA4-1 and shNCOA4-2) subjected to DFO chelation in the presence of lysosomal protease inhibitors for 9 h. Scale bar, 10 μm. Punctate ferritin fraction was quantified from >75 cells per cell line from two independent experiments (number of U2OS cells quantified is as follows: shGFP, 133 cells; shNCOA4-1, 103 cells; and shNCOA4-2, 79 cells). Bars and error bars represent mean values and s.d., respectively. $***P < 0.001$ using a one-sided *t*-test. **d**, IMR-90 cells were treated and immunostained as in **c**. Scale bar, 50 μm. Punctate ferritin fraction was quantified from >25 cells per cell line in two independent experiments and from more than ten microscopy fields (number of IMR-90 cells quantified is as follows: shGFP, 29 cells; shNCOA4-1, 26 cells; and shNCOA4-2, 31 cells). Bars and error bars represent mean values and s.d., respectively. $***P < 0.001$ using a one-sided *t*-test. **e**, HPDE cells were treated and immunostained as in **c**. Scale bar, 10 μm. Quantification was not possible due to the high background signal in shGFP control cells.



Extended Data Figure 9 | A non-degradable murine NCOA4 rescues RNAi-mediated NCOA4 knockdown. **a**, PA-TU-8988T cells stably expressing either a control MSCV empty vector or the murine homologue of NCOA4 (selected with blasticidin) as well as stably expressing shRNAs (shGFP, shNCOA4-1, shNCOA4-2, selected with puromycin) were cultured in the presence or absence of FAC for 24 h, washed, and subjected to DFO chelation in the presence of lysosomal protease inhibitors (E-64d and pepstatin A). Cells were fixed and immunostained using antibodies to ferritin (red) and LAMP2 (green). Scale bar, 10 µm. Punctate ferritin fraction was quantified from ≥100 cells per cell line (number of PA-TU-8988T cells quantified is as follows: PA-TU-8988T-control MSCV-shGFP, 100 cells; PA-TU-8988T-control-MSCV-shNCOA4-1, 125 cells; PA-TU-8988T-control-MSCV-shNCOA4-2, 132 cells; PA-TU-8988T-mouse-NCOA4-shGFP, 151 cells; PA-TU-8988T-mouse-NCOA4-shNCOA4-1, 153 cells; and PA-TU-8988T-mouse-NCOA4-shNCOA4-2, 172 cells). Bars and error bars represent mean values and s.d., respectively. $***P < 0.001$ using a one-sided *t*-test. **b**, RNAi-mediated knockdown of NCOA4 in PA-TU-8988T cells. PA-TU-8988T cells stably expressing either a control MSCV empty vector or the murine homologue of NCOA4 (selected with blasticidin) as well as stably expressing a control shRNA (shGFP) and two independent shRNAs to NCOA4 (shNCOA4-1 and shNCOA4-2) were lysed and analysed by immunoblotting with an antibody to NCOA4 and ACTB as a loading control. Light and dark exposures are shown. A non-specific band migrates just below the NCOA4-specific band. **c**, Expression of murine HA–NCOA4 protein in PA-TU-8988T cells. Cell lysates were probed with HA antibody.



Extended Data Figure 10 | NCOA4 mediates autophagic delivery of ferritin to lysosomes to control iron homeostasis. **a**, Immunostaining of PA-TU-8988T cells transfected with luciferase control siRNA or two independent siRNAs to NCOA4 and subjected to DFO chelation in the presence of lysosomal protease inhibitors for 9 h. Scale bar, 20 μm. **b**, Immunostaining of PA-TU-8988T cells transfected with luciferase control siRNA or two independent siRNAs to HERC2 and subjected to DFO chelation in the presence of lysosomal protease inhibitors for 9 h. Scale bar, 20 μm. **c**, PA-TU-8988T cells were transfected with luciferase control siRNA or two independent siRNAs to HERC2. Lysates were immunoblotted using antibodies to HERC2 and ACTB. **d**, siRNA-mediated knockdown of NCOA4 in U2OS, IMR-90 and PA-TU-8988T cells leads to increases in IRP2, FTH1 and TFRC levels. Cells were transfected with luciferase control siRNA or two independent siRNAs to NCOA4. Lysates were immunoblotted using antibodies to NCOA4, IRP2, TFRC, FTH1 and ACTB. Light and dark exposures are shown for TFRC. **e**, NCOA4 expression levels are affected by iron levels and NCOA4 levels affect ferritin levels in response to iron load. PA-TU-8988T cells stably expressing either a control MSCV empty vector or mouse NCOA4 were cultured in the presence of FAC for the indicated times. Cells were lysed and analysed by immunoblotting with antibodies to NCOA4, FTH1 and ACTB as a loading control. A non-specific band migrates just below the NCOA4-specific band.

doi:10.1038/nature13140

Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I

Alys Peisley[1,2], Bin Wu[1,2], Hui Xu[3], Zhijian J. Chen[3,4] & Sun Hur[1,2]

Ubiquitin (Ub) has important roles in a wide range of intracellular signalling pathways. In the conventional view, ubiquitin alters the signalling activity of the target protein through covalent modification, but accumulating evidence points to the emerging role of non-covalent interaction between ubiquitin and the target. In the innate immune signalling pathway of a viral RNA sensor, RIG-I, both covalent and non-covalent interactions with K63-linked ubiquitin chains (K63-Ub$_n$) were shown to occur in its signalling domain, a tandem caspase activation and recruitment domain (hereafter referred to as 2CARD)[1,2]. Non-covalent binding of K63-Ub$_n$ to 2CARD induces its tetramer formation, a requirement for downstream signal activation[3]. Here we report the crystal structure of the tetramer of human RIG-I 2CARD bound by three chains of K63-Ub$_2$. 2CARD assembles into a helical tetramer resembling a 'lock-washer', in which the tetrameric surface serves as a signalling platform for recruitment and activation of the downstream signalling molecule, MAVS. Ubiquitin chains are bound along the outer rim of the helical trajectory, bridging adjacent subunits of 2CARD and stabilizing the 2CARD tetramer. The combination of structural and functional analyses reveals that binding avidity dictates the K63-linkage and chain-length specificity of 2CARD, and that covalent ubiquitin conjugation of 2CARD further stabilizes the Ub–2CARD interaction and thus the 2CARD tetramer. Our work provides unique insights into the novel types of ubiquitin-mediated signal-activation mechanism, and previously unexpected synergism between the covalent and non-covalent ubiquitin interaction modes.

Effective immune defence against viral infection is dependent upon efficient detection of pathogens by innate immune receptors. One such receptor is RIG-I, which recognizes RNAs from a broad range of viruses and activates the type I interferon (IFN) signalling pathway[4]. RIG-I consists of an amino (N)-terminal 2CARD, which is important for signal activation, and a central DExH/C motif helicase domain and carboxy (C)-terminal domain (CTD), which together function as a viral RNA recognition unit (Fig. 1a). The crystal structure of RNA-free, full-length RIG-I[5] showed that 2CARD forms an intramolecular interaction with the helicase domain. Upon RNA binding, 2CARD was proposed to be released from the helicase domain[5–7], which would allow its interaction with the N-terminal CARD of the downstream signalling adaptor molecule, MAVS[8–10]. Upon interaction with RIG-I 2CARD, MAVS CARD forms a self-perpetuating filament, which then recruits TRAF2, 3 and 6 to activate downstream IFN-α/β signalling pathways[11].

Accumulating evidence suggests that a simple release of 2CARD is insufficient for the signal activation by RIG-I. It was shown that modification of RIG-I 2CARD with K63-Ub$_n$ by the E3 ligase TRIM25 is important for signalling activity of isolated 2CARD[2]. Subsequent studies showed that unanchored K63-Ub$_n$ can non-covalently bind 2CARD to promote its tetramerization and concomitant signal activation[1,3]. Recently, we and others have shown that RIG-I forms a filament along double-stranded RNA (dsRNA) using its helicase and CTD[12,13], and that the resultant proximity can promote 2CARD oligomerization and stimulate MAVS in the absence of K63-Ub$_n$[12]. This differs from isolated 2CARD, which strictly requires K63-Ub$_n$ for oligomerization and signal activation[1,3] (Extended Data Fig. 1b). These studies suggest the presence of at least three mechanisms, Ub-conjugation, Ub-binding and filament formation, and more complex signal activation process for RIG-I.

To understand more precisely the roles of covalent and non-covalent interactions with Ub, we have crystallized human RIG-I 2CARD in complex with K63-Ub$_2$, a minimal chain length required for tetramerization of 2CARD[1]. Our initial crystallization trials included wild-type and 2CARD surface mutants, but crystals were obtained only with the double mutant, K115A/R117A, which retains the wild type's ability to tetramerize with K63-Ub$_n$ and activate the IFN-β pathway (Extended Data Fig. 1a, b). The structure was determined by molecular replacement and was refined to 3.7Å with R_{work} of 22.2% and R_{free} of 28.5% (Extended Data Table 1). The structure revealed the tetrameric architecture of 2CARD bound with six Ub molecules in the asymmetric unit (Fig. 1a, b and Extended Data Fig. 2a). Despite poor electron density for the K63-linkage, SDS–polyacrylamide gel electrophoresis (SDS–PAGE) analysis of the crystal revealed that K63-Ub$_2$ was intact (Extended Data Fig. 2e). Analysis of the distance between the C terminus and K63 of adjacent Ubs unambiguously indicated the three Ub pairs connected through the K63 linkage (Extended Data Fig. 2f).

The RIG-I 2CARD tetramer is formed by a rigid-body docking of monomeric 2CARD. The four 2CARD subunits (A–D) display a conformation that is indistinguishable from 2CARD in the monomeric RIG-I[5] (Extended Data Fig. 3a) in terms of both individual CARD structures and the relative orientation of the first and second CARDs. Superposition of the full-length RIG-I onto one of the 2CARD subunits in the tetramer showed that a part of the 2CARD–2CARD interface is directly masked by the helicase domain in the full-length RIG-I (Extended Data Fig. 3b), indicating that blockade of the tetramerization surface mediates auto-repression of RNA-free RIG-I.

The RIG-I 2CARD tetramer exhibits a helical assembly, in which the adjacent 2CARD subunits (A–B, B–C and C–D) are related by $\sim$101° rotation and $\sim$5Å rise along the central screw axis (Extended Data Fig. 3c). The interaction between subunits D and A at the helical 'seam' (Fig. 1b) differs from the A–B, B–C and C–D interactions by displacement of subunit A by a single CARD (Extended Data Fig. 3c). Consequently, the 2CARD tetramer can be viewed as a 'lock washer' with the two ends displaced by half the thickness of the ring (Fig. 1b). Alternatively, it can be viewed as two continuous helical turns of a single CARD, first with four copies of the first CARD followed by four copies of the second CARD. The helical assembly of 2CARD cannot extend beyond the tetramer as the assembly unit is tandem CARD domains, in which the first and second CARDs are rigidly joined through tight intramolecular interactions.

CARD belongs to the death domain (DD) superfamily, members of which share little sequence similarity, but have a common three-dimensional fold[14]. The helical assembly of 2CARD is reminiscent of those of DD oligomers, such as the Myddosome[15], PIDDosome[16] and

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115 USA. [2]Program in Cellular and Molecular Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [3]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [4]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.

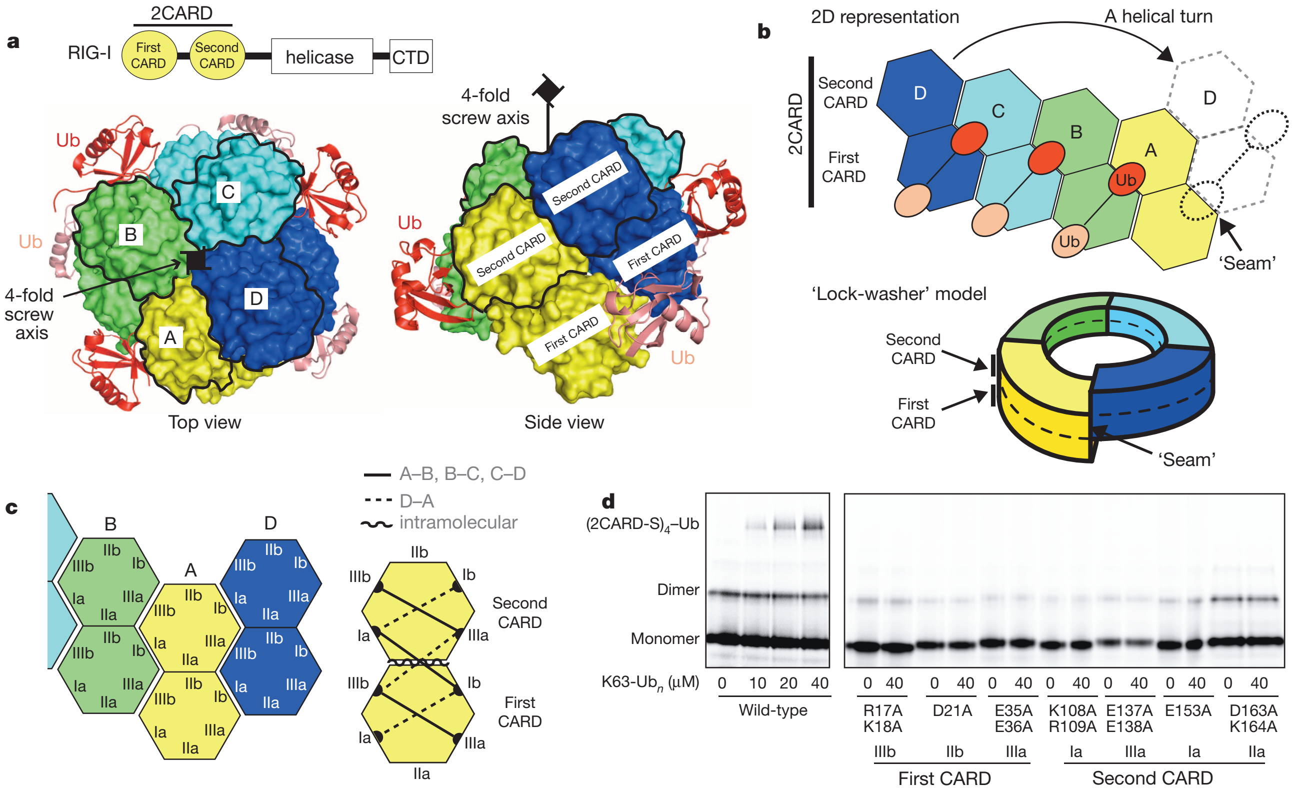


Figure 1 | RIG-I 2CARD assembles into a helical tetramer bridged by three copies of K63-Ub$_2$. **a**, Domain architecture of RIG-I and top and side views of the 2CARD tetramer (chains A–D, surface representation) bound by Ub chains (cartoon representation). The second CARDs are demarcated by black lines. **b**, Two-dimensional (2D) representation and 'lock-washer' model of the RIG-I 2CARD tetramer, using the same colour code as for **a**. **c**, Definition of the surface types in the first and second CARDs. Six types of intermolecular interactions were observed; three in the A–B, B–C and C–D interactions (solid lines on the right), and another three in the D–A interaction (dotted lines). A single type of intramolecular interaction was observed between IIb of the first CARD and IIa of the second CARD (curved line). See Extended Data Fig. 4. **d**, Tetramer formation of wild-type and mutant 2CARD fused to an Alexa647-labelled SNAP tag (2CARD-S) with and without K63-Ub$_n$ ($n > 8$) as measured by EMSA. Mutations are on the indicated surface as defined in **c**.

FAS–FADD complex[17]. Assembly of DD oligomers is commonly mediated by six surface areas (Ia/b, IIa/b and IIIa/b) forming three types of intermolecular interactions (Ia–Ib, IIa–IIb and IIIa–IIIb)[14]. Similarly, RIG-I 2CARD uses the analogous surface areas of both the first and second CARDs for the intramolecular (IIa–IIb) and intermolecular interactions (Ia–Ib and IIIa–IIIb) in the tetramer (Fig. 1c). The six types of Ia–Ib and IIIa–IIIb interactions were observed; three (solid lines in Fig. 1c) between adjacent subunits along the helical trajectory (A–B, B–C and C–D), and the other three (dotted lines) at the helical seam between subunits D and A. Mutation of residues involved in the inter- or intramolecular interactions impaired the ability of 2CARD to tetramerize with K63-Ub$_n$ (as measured by electrophoretic mobility shift assay (EMSA) using 2CARD fused to a fluorescent tag, SNAP[12], Fig. 1d), and to activate the IFN-β signalling pathway in cells (Extended Data Fig. 3d). This result suggests the importance of inter-CARD interactions in formation of the signalling competent RIG-I 2CARD tetramer.

We next asked how this tetrameric, but not monomeric, 2CARD can activate downstream signalling pathway, that is, MAVS CARD filament formation. In the case of Myddosome, PIDDosome, and FAS–FADD complex, the helical oligomeric structure of upstream signalling molecules serves as a scaffold to assemble the downstream DD oligomers through helical extension[14]. We speculated that RIG-I 2CARD may use a similar mechanism to assemble the MAVS CARD filament (Fig. 2a). According to this model, the bottom surface of the first CARD or the top of the second CARD would serve as a platform to recruit MAVS CARD, and the resultant MAVS CARD filament would use the same type of molecular interactions (Ia–Ib, IIa–IIb and IIIa–IIIb) as in the 2CARD tetramer (Fig. 2a). We found that mutations of surface IIb (top) of the second CARD, but not IIa (bottom) of the first CARD, significantly reduced the signalling activity of RIG-I (Fig. 2c), suggesting that the top of the second CARD may be involved in MAVS recruitment. Consistently, mutations in IIb of the second CARD completely abrogated its ability to stimulate MAVS CARD filament formation without affecting its tetramer formation of 2CARD (Fig. 2d, e), suggesting that the top surface of the second CARD serves as the primary site for MAVS recruitment. In addition, both MAVS CARD filament assembly and cellular signalling activity was significantly affected by mutations in the equivalent surface areas, Ia/b-IIIa/b (Fig. 2f, g), further supporting our 'helix-extension' model of MAVS activation by the RIG-I 2CARD tetramer. Although the helical symmetry of RIG-I 2CARD (that is, a pitch in the screw equivalent to the height of a single CARD) prevents helical assembly of RIG-I tandem CARDs beyond the tetramer, it is compatible with filament extension of MAVS as it only has a single CARD.

Unlike the DD oligomers, tetramer formation of isolated 2CARD requires K63-Ub$_n$. In our structure, three chains of K63-Ub$_2$ were bound along the outer rim of the helical trajectory (Fig. 3a). Each chain interacts with two adjacent 2CARDs in the same manner (Extended Data Fig. 5a). There are two types of interactions between Ub and 2CARD; the 'proximal' Ub fits into the groove between the adjacent 2CARDs, forming composite interactions using two distinct surface patches of Ub (L8/I44/V70 and F45/A46/N60/Q62) (Fig. 3a, b). These two patches of proximal Ub contact surface IV and V of 2CARD, respectively. By contrast, the 'distal' Ub interacts with only one 2CARD (surface VI) using the L8/I44/V70 surface patch (Fig. 3a, b). One exception is the distal Ub bound at the helical seam, which interacts with both subunits D (surface VI) and A (surface V). Note that the location of the surface IV in the second CARD is equivalent to that of VI in the first CARD (Extended Data Fig. 5b). Mutations in any one of the three surface areas (IV, V and VI) in 2CARD impaired its ability to tetramerize upon addition of free K63-Ub$_n$ (Fig. 3c). However, the cellular signalling activity

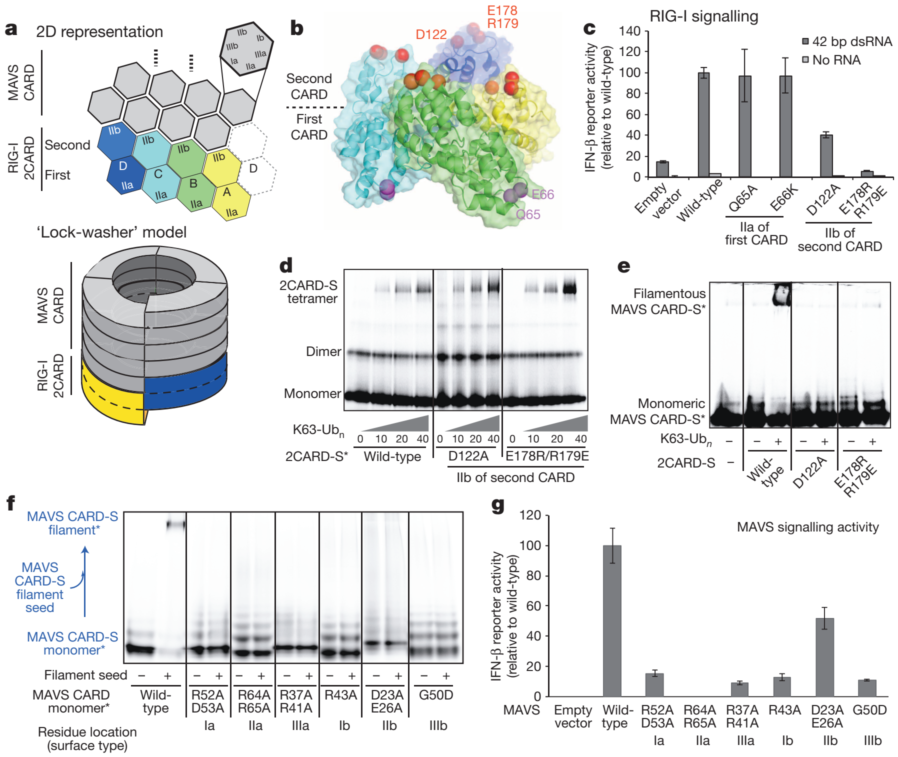


Figure 2 | The helical tetramer of RIG-I 2CARD uses the top surface of the second CARD to interact with MAVS CARD and promote MAVS filament formation. **a**, A 'helical extension' model for how the RIG-I 2CARD tetramer stimulates the MAVS CARD (grey) filament formation. **b**, D122 and E178/R179 are located at the top of the 2CARD tetramer (surface IIb of the second CARD), whereas Q65 and E66 are at the bottom of the 2CARD tetramer (surface IIa of the first CARD). **c**, IFN-β signalling activity of mutants with altered surface IIa or IIb of the first and second CARDs (mean ± s.d., $n = 3$). **d**, **e**, 2CARD tetramerization (**d**) and MAVS filament stimulation (**e**) of wild-type RIG-I 2CARD and its mutants. The SNAP-fusion allowed fluorescent labelling (indicated with an asterisk) of RIG-I 2CARD and MAVS CARD (2CARD-S and CARD-S, respectively). **f**, **g**, Filament formation (**f**) and IFN-β signalling activity (**g**) (mean ± s.d., $n = 3$) of wild-type MAVS and its mutants with altered surface I–IIIa/b.

of RIG-I 2CARD (fused to glutathione *S*-transferase (GST)) was impaired by mutations of surface V and VI, but not IV (Fig. 3d), suggesting that the 2CARD(IV)–Ub interaction is compensated by other factors in cells, such as covalent conjugation of Ub near IV (explained further in Fig. 4).

The observed K63-linkage specificity of 2CARD[3] appears to be dictated by avid binding of K63-Ub$_n$ chains to the proximal and distal Ub binding sites, as observed with other Ub receptors[18]. That is, the relative distance and orientation of the proximal and distal Ub binding sites are compatible with K63-linked, but not K48-linked or linear Ub chains (Extended Data Fig. 5c). The poor electron density for the K63-linker, an indicator of conformational flexibility, also argues against a mechanism in which the K63-linkage is directly recognized by 2CARD, although other factors could have also contributed to conformational heterogeneity of the linker (Extended Data Fig. 2g).

The structure predicts that longer Ub chains could wrap around the 2CARD tetramer at 1:4 or 2:4 molar ratios, thereby increasing the binding avidity and efficiency to promote 2CARD tetramerization. We observed that the efficiency of 2CARD tetramer assembly consistently

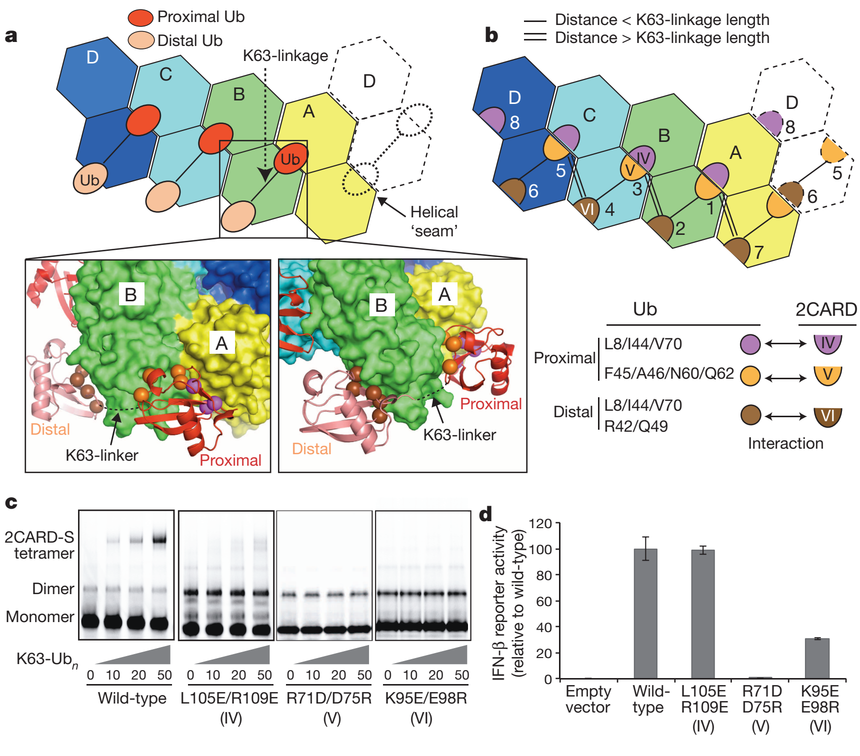


Figure 3 | Ub binds to the 2CARD tetramer by decorating the outer rim of the helical trajectory. **a**, Two types of Ub–2CARD interactions. Proximal Ub forms a composite interaction with two adjacent 2CARDs using two surface areas (purple and orange spheres). Distal Ub interacts with only one 2CARD using one of the two surface areas (brown spheres), with the exception of Ub bound to D. Dotted lines represent disordered K63-linkers between proximal and distal Ubs. See Extended Data Fig. 6. **b**, Three types of 2CARD surface areas (IV, V and VI) are involved in Ub binding. Surface IV and VI interact with L8/I44/V70 in proximal and distal Ubs, respectively, whereas surface V interacts with F45/A46/N60/N62 of proximal Ub. Mapping of IV–VI onto the 2D representation of the 2CARD tetramer revealed 7 bona fide Ub binding sites, six of which (sites 1–6) are occupied in our structure, and one potential binding site (site 8, see Extended Data Fig. 7c, d for additional discussion). Single and double lines indicate respectively the distance between adjacent Ubs that can be directly connected through a K63-linkage and those that cannot (thus requiring a spacer Ub). **c**, EMSA analysis of the tetramer formation by wild-type RIG-I 2CARD-S and its mutants with altered surfaces IV, V and IV, upon incubation with K63-Ub$_n$ ($n > 8$). **d**, IFN-β reporter activity of wild-type and mutant RIG-I 2CARD in GST fusion (mean ± s.d., $n = 3$).

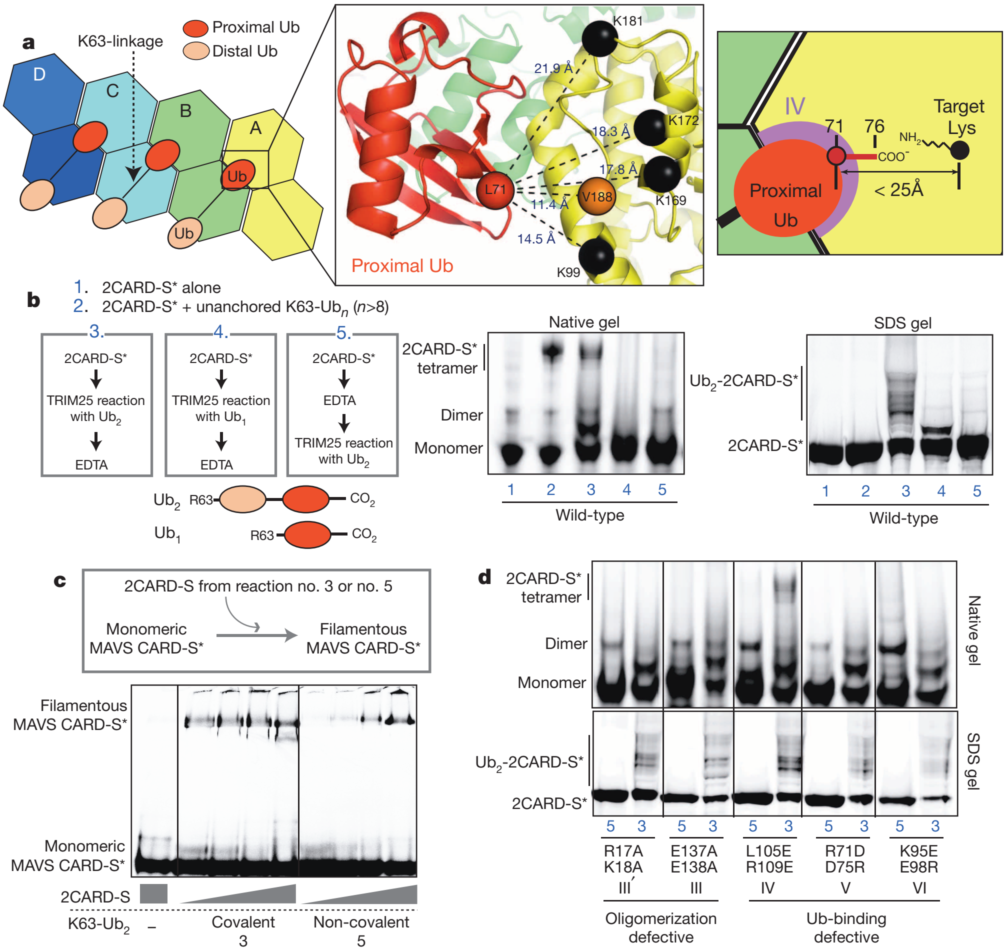


Figure 4 | Covalent conjugation and non-covalent binding of Ub synergize to stabilize the 2CARD tetramer. **a**, The C-terminal residue of the proximal Ub is within the covalent-linkage distance from the target Lys. As residues 72–76 of proximal Ub were disordered in the structure, distance was measured between Cαs of L71 in proximal Ub (red sphere) and target Lys (K99, K169, K172 and K181, black spheres) in 2CARD. V188 was used in place of K190 and K193, which are in the disordered C-terminal tail of 2CARD. The distance requirement for covalent conjugation (<25Å) is shown on the right. **b**, EMSA analysis of 2CARD tetramerization stimulated by covalently conjugated or unanchored K63-Ub$_2$ or Ub$_1$. Fluorescently labelled (indicated with an asterisk) 2CARD-S was subject to Ub conjugation (TRIM25 reaction) before or after EDTA quenching. Samples 3 and 5 enable direct comparison between covalently conjugated and unanchored Ub$_2$. As elongation of Ub chains by TRIM25 is prevented by K63R mutation within Ub$_2$ and Ub$_1$, the presence of several bands in SDS–PAGE analysis reflects conjugation of Ub$_2$ to 2CARD at multiple Lys targets. **c**, Comparison of the MAVS stimulatory activity of 2CARD covalently conjugated with K63-Ub$_2$ (sample 3 from **b**) or 2CARD with unanchored K63-Ub$_2$ (sample 5 from **b**). Increasing concentrations of 2CARD-S (47 or 187 or 750 or 3000 nM) were incubated with fluorescently labelled monomeric MAVS CARD-S and MAVS filament formation was monitored by EMSA. **d**, EMSA analysis of tetramerization of mutant 2CARD by covalently conjugated (sample 3 from (**b**)) or unanchored K63-Ub$_2$ (sample 5 from (**b**)).

increases with Ub chain length and that a single chain of Ub$_3$ can stabilize the 2CARD tetramer (Extended Data Fig. 7a, b, g). To understand how K63-Ub$_n$ with $n > 2$ interacts with the 2CARD tetramer, we identified possible Ub-binding sites by mapping the IV, V and VI sites onto the 2CARD tetramer (Fig. 3b). Besides the six sites occupied in our crystal structure (sites 1–6 in Fig. 3b), we located two additional Ub binding sites (site 7, and possibly 8) (Extended Data Fig. 7c, d). As the distance between sites 2 and 3; 4 and 5; and 7 and 1 is too great to be directly connected via the K63 linkage (Fig. 3b), a 'spacer' Ub would be required between these sites (Extended Data Fig. 7e). Accordingly, the chain length of $n = 5$–10 would be required to bridge the four 2CARD subunits (Extended Data Fig. 7e, f), which coincides with the chain length of K63-Ub$_n$ co-purified with RIG-I 2CARD from cells[1].

The relationship between the covalent conjugation and non-covalent binding of K63-Ub$_n$ in RIG-I signalling has been controversial[1,2]. These two mechanisms have been proposed to use the same six Lys residues (K99, K169, K172, K181, K190 and K193)[1,2]. Our structure showed that none of these residues are involved in K63-Ub$_n$ binding or in the 2CARD–2CARD interaction, suggesting that the observed negative impact of mutating the target lysines (6KR)[2] reflects the importance of covalent Ub conjugation in the signalling activity of isolated 2CARD (Extended Data Fig. 5d). To understand the interplay of covalent and non-covalent binding of Ub, we examined the distance between the target lysine and the C-terminal tail of proximal Ub. Although 2CARD residues 189–200 containing K190 and K193 are disordered in the crystal structure, the remaining four target lysines and the most C-terminal residue, V188, are within the covalent linkage distance from the C-terminal tail of proximal Ub (Fig. 4a). This observation suggests that covalent conjugation of K63-Ub$_n$ to the six lysines is compatible with its non-covalent binding to 2CARD, and may further stabilize the 2CARD tetramer.

To test this hypothesis, we compared the tetramerization efficiency of 2CARD conjugated with K63-Ub$_2$ and 2CARD with free K63-Ub$_2$. We observed the mobility shift of 2CARD when K63-Ub$_2$ was covalently conjugated by TRIM25 (sample 3 in Fig. 4b) similar to the shift observed with unanchored K63-Ub$_n$ ($n > 8$) (sample 2). No shift was observed when mono-Ub was conjugated (sample 4) or Ub$_2$-conjugation was inhibited by EDTA (sample 5), suggesting that the mobility shift was caused by covalent conjugation with Ub$_2$, not by free Ub$_2$ (which requires higher

concentrations to induce 2CARD tetramerization, Extended Data Fig. 7a). Accordingly, Ub_2-conjugated 2CARD stimulated MAVS filament formation more efficiently than 2CARD with free Ub_2 (Fig. 4c). In addition, mutations in the 2CARD–2CARD and 2CARD–Ub interfaces (except for the mutation in IV) abrogated tetramerization of 2CARD conjugated with Ub_2 (Fig. 4d). These results suggest that 2CARD tetramerization mediated by covalently conjugated Ub_2 requires the same type of non-covalent 2CARD–Ub and 2CARD–2CARD interactions as with that mediated by unanchored K63-Ub_n. The loss of tetramerization of Ub_2-conjugated 2CARD upon VI mutation is consistent with the observation that conjugation with mono-Ub (and its bridging of IV and V) is insufficient to induce 2CARD tetramerization (Fig. 4b). As for the IV mutation, covalent conjugation of Ub_2 compensated for inefficient Ub binding (Figs 3c and 4d), which provides a rationale for its robust signalling activity in cells (Fig. 3d). This result further supports the notion that covalent Ub conjugation near its binding site enhances the stability of the 2CARD tetramer.

We have previously shown that induced proximity of 2CARD within RIG-I filaments formed on long (~60–100 base pair (bp)) dsRNA could contribute to 2CARD oligomerization independtly of Ub[12]. To better understand how filament formation interplays with Ub-mediated 2CARD oligomerization, we investigated the dependence of the signalling activity of Ub conjugation-defective (6KR) or binding-defective (K95E/E98R) mutants on RNA length. Both mutations significantly reduced the signalling activity of full-length RIG-I with 21 bp dsRNA (Extended Data Fig. 5d), a dsRNA length that does not support filament formation. However, their negative effects were progressively alleviated by increasing the length of dsRNA (Extended Data Fig. 5d). This result suggests that the requirement for Ub-conjugation and Ub-binding can be partially compensated by filament formation[12], and that the three mechanisms, Ub-binding, conjugation and filament formation, act synergistically for optimal signal activation.

Ub-dependent signal activation typically involves covalent conjugation of a target protein with Ub upon signal activation, and this 'Ub signal' is decoded by a downstream protein which binds to the conjugated Ub in linkage- and length-specific manners[18]. In the RIG-I signalling mechanism, 2CARD is both an Ub-conjugation target and a receptor, with little structural similarity to previously characterized Ub-binding domains. Our current work shows how these dual functions of RIG-I 2CARD synergize and lead to efficient formation of the signalling competent tetramer and provides unique insights into a novel type of Ub-mediated signal activation mechanism. Furthermore, the resultant helical architecture of the RIG-I 2CARD tetramer offers a previously unexpected signal activation mechanism for RIG-I, in which the 2CARD tetramer is used as a signalling 'platform' to recruit and assemble the MAVS filament. Whether there is potential for generalizing our findings to MDA5 or other CARD domains will require future investigations (Extended Data Figs 8 and 9).

METHODS SUMMARY

Human RIG-I 2CARD (K115A/R117A, residue 1–200) was expressed in BL21(DE3) and purified by Ni-NTA affinity and size-exclusion chromatography (SEC). The RIG-I 2CARD–K63-Ub_2 complex was assembled by mixing RIG-I 2CARD and K63-Ub_2 at a 1:2 molar ratio, purified by SEC and crystallized in reservoir buffer containing 0.2 M tri-lithium citrate, 22–24% PEG300 and 3% ethylene glycol. Diffraction data were collected at beamline 8.3.1 at the Advanced Light Source and beamline 24ID-C at the Advanced Photon Source and processed using the program XDS[19]. The structure was determined by molecular replacement using Phaser[20] with duck 2CARD and monoubiquitin (from Protein Data Bank (PDB): 4A2W and 1UBQ, respectively), and refined using Coot[21] and Phenix Refine[22]. A summary of data collection and structure refinement statistics is provided in Extended Data Table 1.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 November 2013; accepted 11 February 2014.

Published online 2 March 2014.

1. Zeng, W. *et al.* Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. *Cell* **141,** 315–330 (2010).
2. Gack, M. U. *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446,** 916–920 (2007).
3. Jiang, X. *et al.* Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity* **36,** 959–973 (2012).
4. Kato, H., Takahasi, K. & Fujita, T. RIG-I-like receptors: cytoplasmic sensors for non-self RNA. *Immunol. Rev.* **243,** 91–98 (2011).
5. Kowalinski, E. *et al.* Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147,** 423–435 (2011).
6. Jiang, F. *et al.* Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* **479,** 423–427 (2011).
7. Luo, D. *et al.* Structural insights into RNA recognition by RIG-I. *Cell* **147,** 409–422 (2011).
8. Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-κB and IRF3. *Cell* **122,** 669–682 (2005).
9. Xu, L. G. *et al.* VISA is an adapter protein required for virus-triggered IFN-β Signaling. *Mol. Cell* **19,** 727–740 (2005).
10. Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6,** 981–988 (2005).
11. Hou, F. *et al.* MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* **146,** 448–461 (2011).
12. Peisley, A., Wu, B., Yao, H., Walz, T. & Hur, S. RIG-I forms signaling-competent filaments in an ATP-dependent, ubiquitin-independent manner. *Mol. Cell* **51,** 573–583 (2013).
13. Patel, J. R. *et al.* ATPase-driven oligomerization of RIG-I on RNA allows optimal activation of type-I interferon. *EMBO Rep.* **14,** 780–787 (2013).
14. Ferrao, R. & Wu, H. Helical assembly in the death domain superfamily. *Curr. Opin. Struct. Biol.* **22,** 241–247 (2012).
15. Lin, S.-C., Lo, Y.-C. & Wu, H. Helical assembly in the MyD88–IRAK4–IRAK2 complex in TLR/IL-1R signalling. *Nature* **465,** 885–890 (2010).
16. Park, H. H. *et al.* Death domain assembly mechanism revealed by crystal structure of the oligomeric PIDDosome core complex. *Cell* **128,** 533–546 (2007).
17. Wang, L. *et al.* The Fas–FADD death domain complex structure reveals the basis of DISC assembly and disease mutations. *Nature Struct. Mol. Biol.* **17,** 1324–1329 (2010).
18. Husnjak, K. & Dikic, I. Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu. Rev. Biochem.* **81,** 291–322 (2012).
19. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
20. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
21. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr.* **60,** 2126–2132 (2004).
22. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).

Acknowledgements We thank G. Meigs at ALS BL 8.3.1 for data collection, and H.Wu and M.Eck for comments on the manuscript. This work was supported by GSK fellowship (B.W.), The Welch Foundation (I-1389; Z.J.C.), NIH (R01-GM63692; Z.J.C), Pew Scholarship (S.H.) and Career Development Award from BCH (S.H.).

Author Contributions S.H. initiated the project idea. A.P., B.W. and S.H. designed the experiments. A.P. and B.W. performed the experiments. H.X. and Z.J.C. provided K63-Ub_{3-8}. A.P. and S.H. interpreted the data and wrote the manuscript.

Author Information The atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession number 4NQK. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H. (Sun.Hur@childrens.harvard.edu).

METHODS

Plasmid constructs. Wild type and variants of RIG-I 2CARD (residue 1–200) were subcloned between XmaI and HindIII restriction sites in pET47b (Novagen). Site-directed mutagenesis was performed using the KAPA HiFi PCR kits (KAPA biosystems). For the 2CARD-SNAP (2CARD-S) construct, 2CARD was subcloned between the XmaI and HindIII, and SNAP between the XhoI and AvrII restriction sites in pET47b. For MAVS CARD fused to SNAP (CARD-S), CARD and SNAP were sequentially inserted between the BamHI and EcoRI, and between the EcoRI and XhoI restriction sites in pET47b, respectively. The bacterial expression construct for mouse E1 ubiquitin activating enzyme (pET28-mE1) was purchased from Addgene. Human TRIM25 was cloned between KpnI and SalI restriction sites of pET50b. Human UbcH5, Ubc13 and Uev1a were inserted between the XmaI and XhoI restriction sites in pET47b, and human ubiquitin was inserted between the XmaI and HindIII restriction sites in pET47b. For expression of RIG-I in HEK293T cells, wild-type RIG-I and its mutants were inserted between the HindIII and KpnI restriction sites in pFLAG-CMV4 (Sigma). GST fusion constructs for RIG-I 2CARD (in pEBG) and its Lys variant (6KR) were gifts from M. U. Gack (Harvard Medical School).

Material preparation. RIG-I 2CARD and 2CARD-S were expressed in BL21(DE3) at 20 °C for 16–20 h following induction with 0.4 mM IPTG. The proteins were purified by Ni-NTA affinity chromatography. The 6×His tag was removed by HRV 3C protease, and 2CARD was further purified by additional Ni-NTA affinity chromatography (which removes uncleaved protein and protease) and SEC in buffer A (20 mM HEPES pH 7.5, 150 mM NaCl). For N-terminal fluorescent labelling of RIG-I 2CARD, the protein (~2 mg ml^{-1}) was incubated with 0.5 mM peptide (LPETGG) conjugated with fluorescein (Anaspec) and 0.3 mM *S. aureus* sortase A (a gift from Hidde Ploegh, MIT)[23] at room temperature (RT) for 2–3 h, followed by Ni-NTA affinity purification to remove sortase and SEC.

MAVS CARD was expressed as a fusion protein with the SNAP tag (CARD-S) in BL21(DE3) at 20 °C for 16–20 h following induction with 0.4 mM IPTG. MAVS CARD-S was purified using Ni-NTA affinity chromatography and SEC. Purified CARD-S exists in the form of short filaments and was denatured in 6 M guanidinium hydrochloride for 30 min at 37 °C, followed by dialysis against 20 mM HEPES, pH 7.5, 500 mM NaCl, 0.5 mM EDTA and 10 mM BME at 4 °C for 1 h. Refolded CARD-S was labelled with Alexa488-benzylguanine (NEB) according to the manufacturer's instruction, further purified by SEC in buffer B (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM EDTA) to remove unincorporated dye, and was immediately used for filament formation assays.

Mouse E1, human UbcH5, Ubc13, Uev1a and ubiquitin were prepared as previously reported[12]. NusA fusion of TRIM25 was expressed from BL21(DE3) at 20 °C for 16–20 h following induction with 0.4 mM IPTG and purified by Ni-NTA affinity and SEC.

The 42 and 112 bp dsRNAs were synthesized as before (ref. 12). See ref. 12 for the sequence.

Synthesis of unanchored and 2CARD-conjugated polyubiquitin chains. Lys63-linked ubiquitin chains (K63-Ub$_n$) were generated from a reaction containing 0.4 mM ubiquitin, 4 μM mE1, 20 μM Ubc13 and 20 μM Uev1a in buffer (10 mM ATP, 50 mM Tris pH 7.5, 10 mM MgCl2, 0.6 mM DTT). The K63-Ub$_n$ synthesis reaction was performed overnight at 37 °C. Synthesized K63-Ub$_n$ chains were purified as described previously[24]. Briefly, ubiquitin chains were diluted fivefold into 50 mM ammonium acetate, pH 4.5, 0.1 M NaCl and separated over a 45 ml 0.1–0.6 M NaCl gradient in 50 mM ammonium acetate, pH 4.5 using a Hi-Trap SP FF column (GE Healthcare). High molecular weight fractions were applied to an S200 10/300 column equilibrated in Buffer A.

Ub$_2$ was generated from a reaction containing 0.2 mM Ub(D77), 0.2 mM Ub(K63R), 4 μM mE1, 20 μM Ubc13 and 20 μM Uev1a in buffer (10 mM ATP, 50 mM Tris pH 7.5, 10 mM $MgCl_2$, 0.6 mM DTT). Ub$_4$, Ub$_6$ and Ub$_8$ were synthesized as previously reported (ref. 1).

For synthesis of Ub$_2$ covalently conjugated to RIG-I 2CARD, 10 μM wild-type and mutant RIG-I 2CARD-S constructs were incubated with 350 μM Ub$_2$, 4 μM mE1, 2 μM TRIM-25, 150 μM UbcH5 in 50 mM Tris, pH 7.5, 10 mM ATP, 10 mM $MgCl_2$, 2 mM DTT for 2 h at 37 °C.

Crystallization and structure determination. The RIG-I 2CARD–K63-Ub$_2$ complex was assembled by mixing RIG-I 2CARD (residues 1–200, K115A/R117A) (~10.0 mg ml^{-1}) and K63-Ub$_2$ at a 1:2 molar ratio, and was concentrated to ~10 mg ml^{-1} in buffer A and 2 mM DTT and the resultant complex was purified by SEC. Crystals were initially obtained as needle clusters using the hanging-drop vapour-diffusion method from a 1:1 mixture of sample and reservoir buffer that contained 0.2 M tri-lithium citrate, 20% PEG3350. Crystals were optimized using reservoir buffer containing 0.2 M tri-lithium citrate, 22–24% PEG300, 3% ethylene glycol. Diffraction data were collected at beamline 8.3.1 at the Advanced Light Source and beamline 24ID-C at the Advanced Photon Source and processed using the program XDS[19].

The structure was determined by molecular replacement using Phaser[20] with duck 2CARD and monoubiquitin (from PDB: 4A2W and 1UBQ, respectively). We first identified the molecular replacement solution for four copies of 2CARD in the asymmetric unit, which was then fixed in the subsequent rounds of molecular replacement to locate six copies of Ub. The solutions for four copies of 2CARD in the tetramer (chains A through D) were unambiguously identified with the translational function Z-scores (TFZ) between 13 and 24. The solutions for Ub chains E, F, G, H and J were also unambiguous with TFZ between 11 and 23. Two competing solutions for the Ub chain I (Ia and Ib in Extended Data Fig. 2b) were found during earlier stages of molecular replacement with TFZs of 15 and 19, respectively. However, once the rest of the complex (that is, four 2CARDs and Ub chains E-H and J) was fixed as a known solution, only Ia was identified. In addition, only Ia, but not Ib, is compatible with the simulated annealing omit map density (Extended Data Fig. 2b). These results suggest that Ia is the correct solution and thus was selected in the final model. The structure was refined by an iterative process of manual model building using Coot[21] and simulated annealing, individual site, TLS refinement with torsion NCS and bulk solvent scaling using Phenix Refine[22]. Individual CARDs (first and second) and Ub moieties were used as independent TLS groups. The quality of the final model was examined using MolProbity[25] and simulated annealing omit maps (Extended Data Fig. 2a, b) calculated with Phenix. Ramachandran plot indicated 94.3% and 0.9% of the protein residues are in favoured and outlier conformations, respectively. Unless mentioned otherwise, molecular graphics figures were generated using PyMOL (Schrodinger). A summary of data collection and structure refinement statistics is provided in Extended Data Table 1.

2CARD tetramerization assay (EMSA). Unless stated otherwise, EMSA was performed by Alexa647 labelled, purified RIG-I 2CARD-S (10 μM or indicated amount) in buffer A (20 mM HEPES, pH 7.5, 150 mM NaCl, and 2 mM DTT) in the presence and absence of indicated amount of K63-Ub$_n$ (see Extended Data Fig. 1b) for 15 min at room temperature, and the complex was analysed on Bis-Tris native PAGE (Life). Fluorescent gel images were recorded using an FLA9000 scanner (Fuji) and analysed with Multigauge (Fuji).

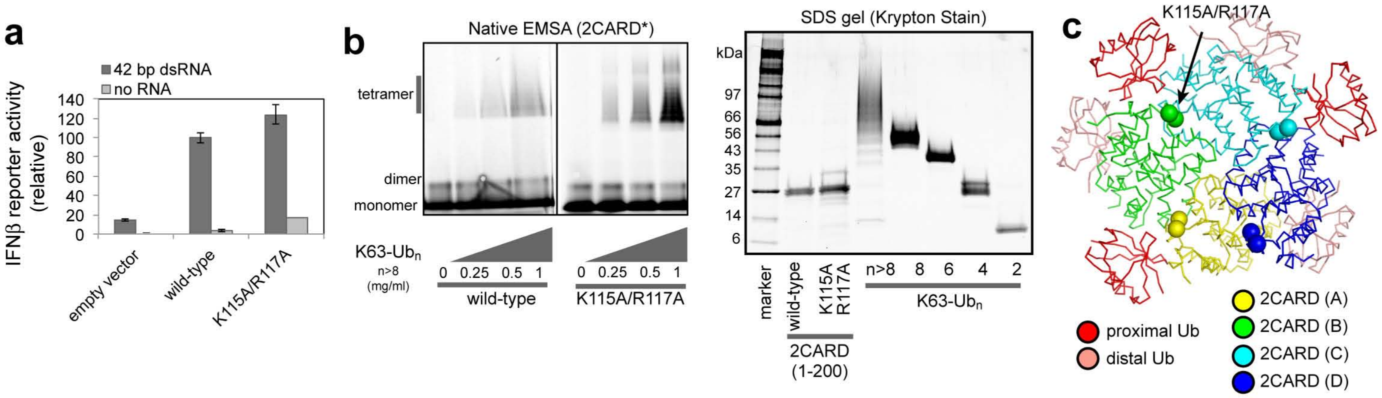
MAVS filament formation assay. The MAVS filament formation assay was performed as previously reported[26]. Refolded, Alexa647 labelled monomer of CARD-S was prepared as previously described[26]. In the absence of external stimuli or seed filaments, refolded MAVS CARD remains stably as a monomer over 24 h, after which it spontaneously forms prion-like filaments over the course of days. Thus, all assays involving MAVS filament formation were performed within 6 h after refolding. To monitor stimulation of MAVS filament formation by RIG-I 2CARD, labelled, monomeric MAVS CARD-S (CARD-S*, 10 μM) was incubated with 10 μM 2CARD-S and 10 μM Ub$_n$ (10 μM monomeric Ub concentration) for 15 min at room temperature before analysis by Bis-Tris native PAGE (Life) or by electron microscopy. For MAVS filament extension assay in Fig. 2f, monomeric CARD-S* was mixed with unlabelled, wild-type MAVS CARD-S filament seeds, and filament extension by the monomeric CARD-S* was monitored by EMSA. Fluorescent gel images were recorded using an FLA9000 scanner (Fuji). Some level of divergence observed between *in vitro* MAVS filament formation assay and cellular IFN-β reporter assay may reflect differences in the sensitivity of the assays and signal readout mechanisms.

Interferon reporter assay. 293T cells were plated in 48-well plates in Dulbecco's modified Eagle medium (Cellgro) supplemented with 10% heat-inactivated fetal calf serum (Gibco) and 1% penicillin/streptomycin. At ~95% confluence, cells were co-transfected with pFLAG-CMV4 plasmids encoding RIG-I or mutants (20 ng), IFN-β promoter driven firefly luciferase reporter plasmid (100 ng) and a constitutively expressed Renilla luciferase reporter plasmid (pRL-TK, 10 ng) by using at 1:2 ratio of DNA:lipofectamine2000 (Life) according to the manufacturer's protocol. The medium was changed 4–6 h post-transfection and cells were subsequently transfected with *in vitro* transcribed dsRNA (0.5 μg) using 1:1 ratio of lipofectamine2000. For MAVS and GST–2CARD, cells were transfected with plasmids expressing full-length MAVS (MAVS-pcDNA3, 20 ng) or GST–2CARD (2CARD-pEBG, 30 ng) with luciferase reporter plasmids (as above) without additional stimulation with dsRNA. Cells were lysed 24 h post-stimulation and IFN-β promoter activity was measured using the Dual Luciferase Reporter assay (Promega) and a Synergy2 plate reader (BioTek). Firefly luciferase activity was normalized against Renilla luciferase activity. Error bars represent standard deviation of three to five independent experiments.

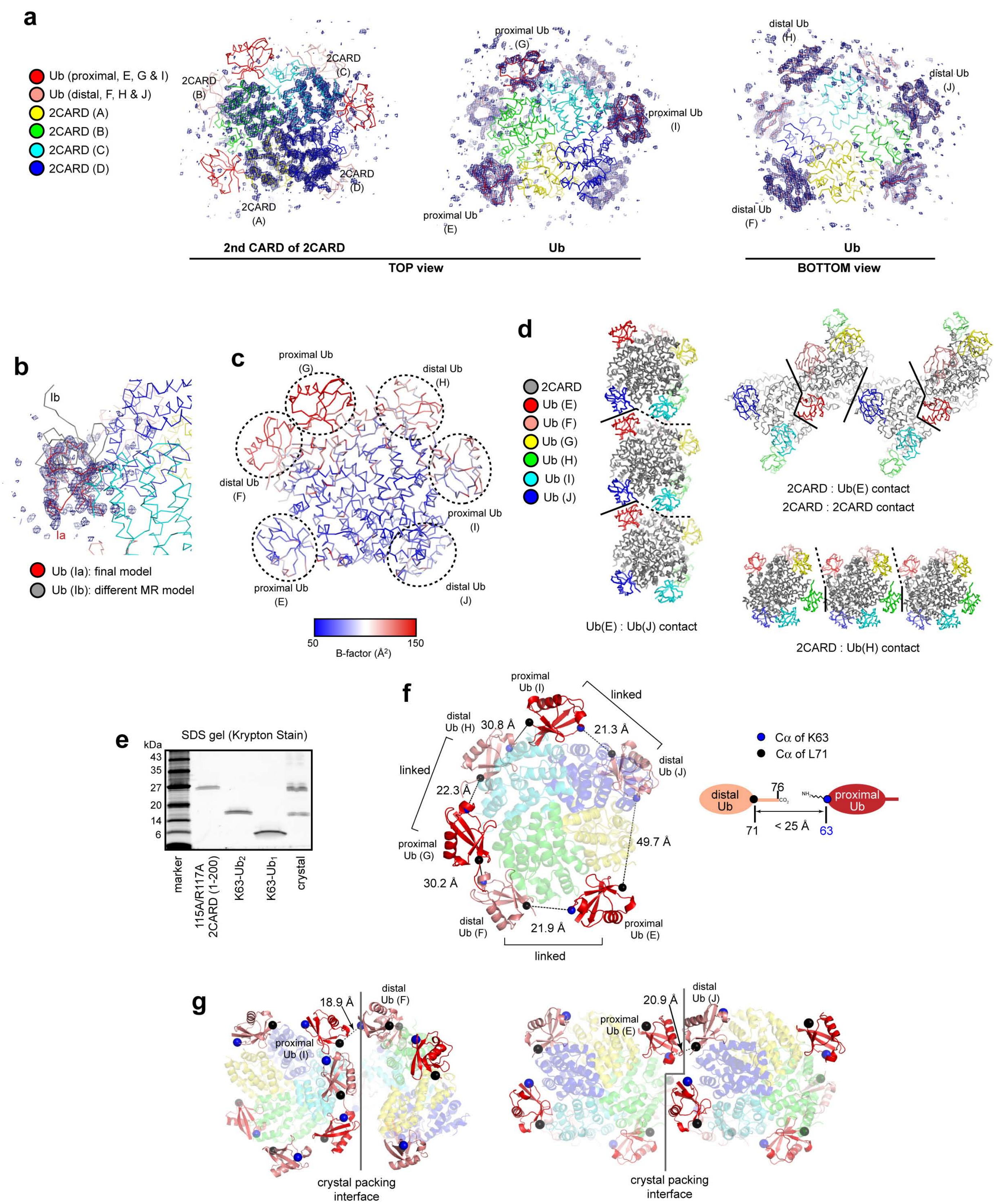
Multi-angle light scattering (MALS). The molecular masses of wild-type RIG-I 2CARD (residues 1–200), K63-Ub$_3$ and that of their complex were determined by MALS using a Superose 200 10/300 column (GE) attached to a miniDAWN TRI-STAR detector (Wyatt Technology) in 20 mM HEPES, pH 7.5, 150 mM NaCl. The complex of 2CARD and K63-Ub$_3$ was formed by mixing at 1:2 molar ratio of 2CARD–K63-Ub$_3$ before loading on a gel filtration column.

23. Antos, J. M. *et al.* Site-specific N- and C-terminal labeling of a single polypeptide using sortase of different specificity. *J. Am. Chem. Soc.* **131,** 10800–10801 (2009).

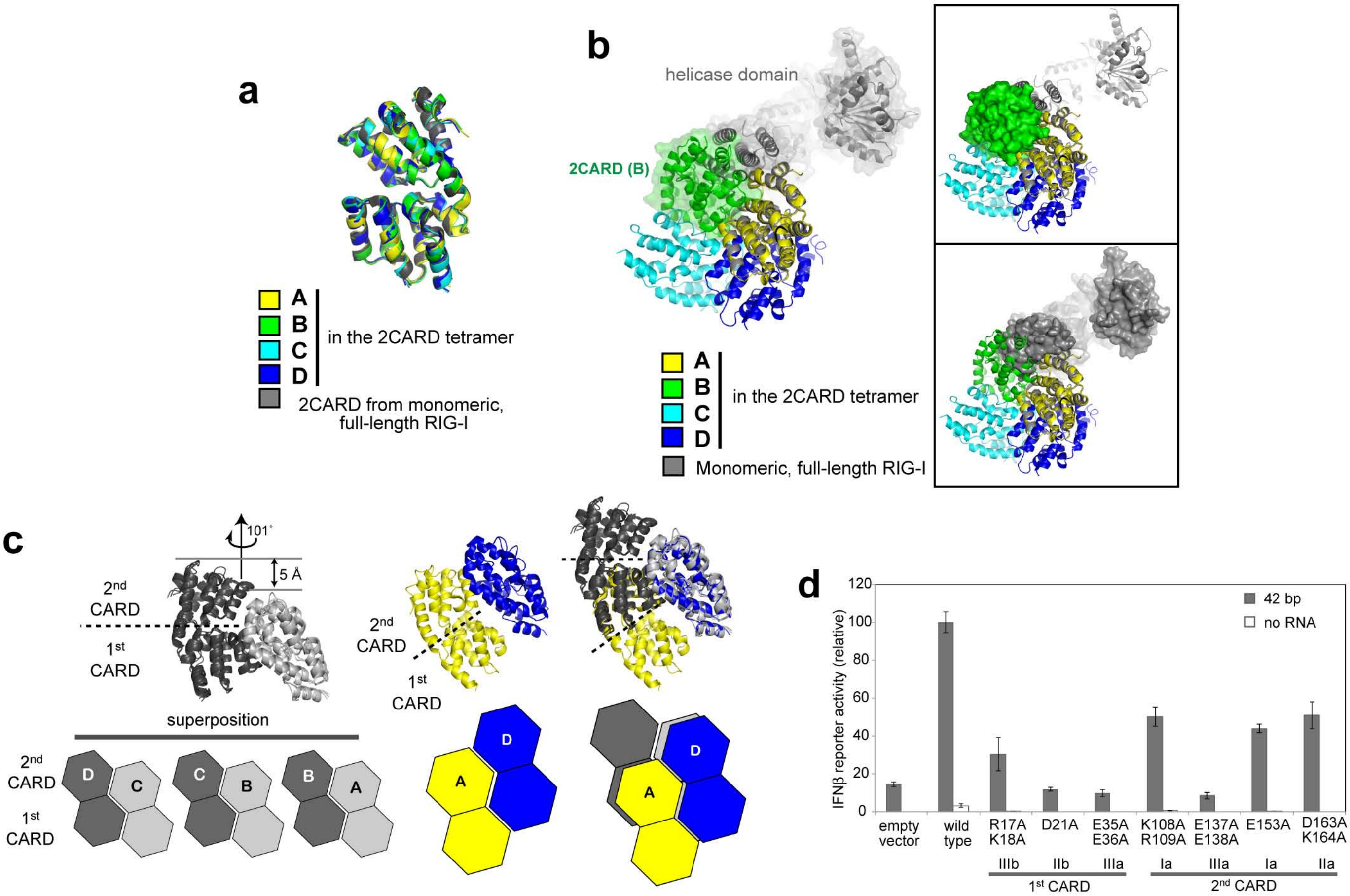
24. Dong, K. C. *et al.* Preparation of distinct ubiquitin chain reagents of high purity and yield. *Structure* **19,** 1053–1063 (2011).
25. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
26. Wu, B. *et al.* Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. *Cell* **152,** 276–289 (2013).
27. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
28. Caffrey, D. R., Somaroo, S., Huges, J. D., Minstseris, J. & Huang, E. Are protein-protein interfaces more conserved sequence than the rest of the protein surface? *Protein Sci.* **13,** 190–202 (2004).
29. Andreani, J., Faure, G. & Guerois, R. Versatility and invariance in the evolution of homologous heteromeric interfaces. *PLOS Comput. Biol.* **8,** e1002677 (2012).
30. Braberg, H. *et al.* SALIGN: A webserver for alignment of multiple protein sequences and structures. *Bioinformatics* **28,** 2072–2073 (2012).



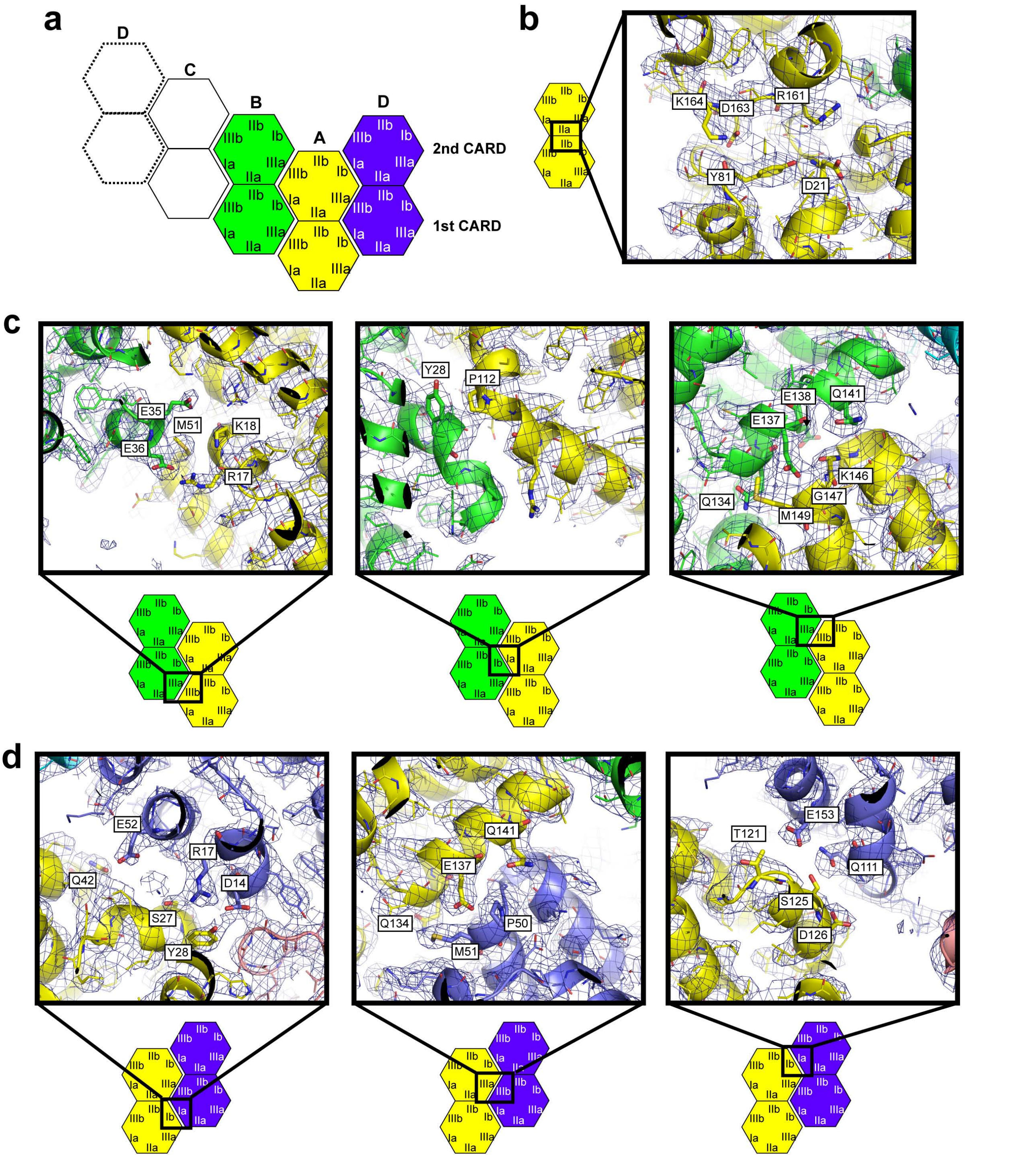
Extended Data Figure 1 | RIG-I 2CARD (K115A/R117A) forms the signalling-competent 2CARD tetramer. **a**, IFN-β reporter activity of wild-type RIG-I and the K115A/R117A mutant with and without 42 bp dsRNA stimulation (mean ± s.d., $n = 3$). **b**, Left, EMSA analysis of tetramerization of wild-type and K115A/R117A RIG-I 2CARD (residues 1–200) with K63-Ub$_n$ ($n > 8$). 2CARD was N-terminally labelled with fluorescein using sortase (see Methods). Right, SDS analysis of wild-type and mutant 2CARD, and K63-Ub$_n$ chains used in this study. Unless mentioned otherwise, K63-Ub$_n$ indicates the chain length $n > 8$ throughout the manuscript. **c**, Mapping of K115 and R117 onto the crystal structure. Although K115 and R117 are located at the edge of the interface, K115A/R117A has little effect on the cellular signalling activity of RIG-I (**a**) and tetramerization of RIG-I 2CARD (**b**), indicating that these residues are not critical for mediating inter-domain contacts.


a
Ub (proximal, E, G & I)
Ub (distal, F, H & J)
2CARD (A)
2CARD (B)
2CARD (C)
2CARD (D)
2CARD (B)
2CARD (C)
2CARD (D)
2CARD (A)
proximal Ub (G)
proximal Ub (I)
proximal Ub (E)
distal Ub (H)
distal Ub (J)
distal Ub (F)
2nd CARD of 2CARD
Ub
TOP view
Ub
BOTTOM view
b
Ib
Ia
Ub (Ia): final model
Ub (Ib): different MR model
c
proximal Ub (G)
distal Ub (H)
distal Ub (F)
proximal Ub (I)
proximal Ub (E)
distal Ub (J)
50
150
B-factor (Å²)
d
2CARD
Ub (E)
Ub (F)
Ub (G)
Ub (H)
Ub (I)
Ub (J)
Ub(E) : Ub(J) contact
2CARD : Ub(E) contact
2CARD : 2CARD contact
2CARD : Ub(H) contact
e
SDS gel (Krypton Stain)
kDa
43
35
27
20
14
6
marker
115A/R117A 2CARD (1-200)
K63-Ub2
K63-Ub1
crystal
f
proximal Ub (I)
linked
distal Ub (H)
30.8 Å
21.3 Å
distal Ub (J)
linked
22.3 Å
proximal Ub (G)
49.7 Å
30.2 Å
distal Ub (F)
21.9 Å
proximal Ub (E)
linked
Cα of K63
Cα of L71
distal Ub
76
proximal Ub
71
< 25 Å
63
g
18.9 Å
distal Ub (F)
proximal Ub (I)
crystal packing interface
20.9 Å
distal Ub (J)
proximal Ub (E)
crystal packing interface

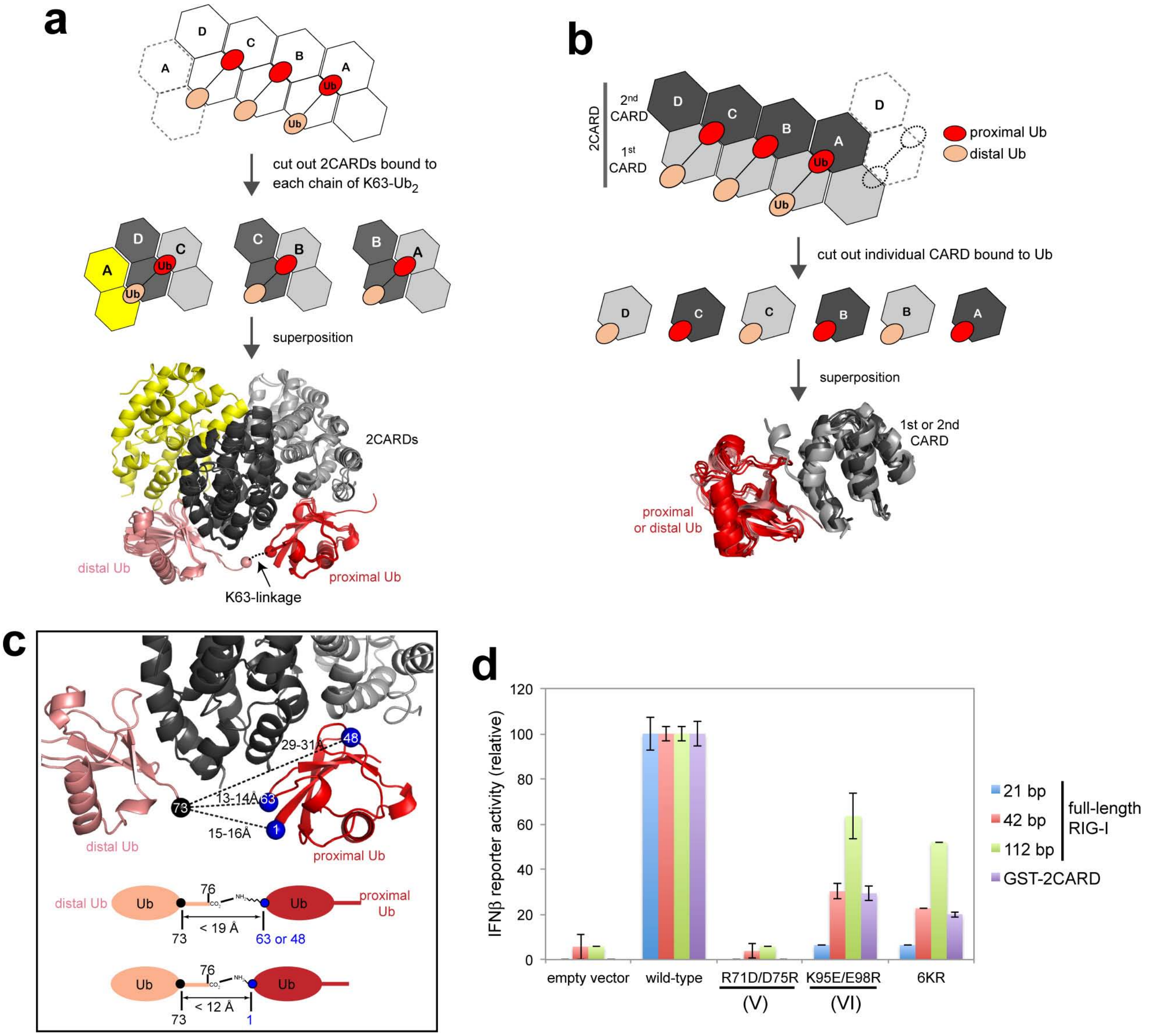
Extended Data Figure 2 | Analysis of the crystal structure of the 2CARD–Ub complex. **a**, Simulated annealing omit maps (sigma-A weighted F_o-F_c, contoured at $\sigma = 2.5$). Four and six maps were separately calculated using models with individual second CARDs (left) and Ubs (middle and right) omitted, respectively, and were overlaid on the crystal structure. The omit maps for Ubs are shown from both the top (middle) and bottom (right) views for better visualization of both proximal and distal Ubs. Although the omit maps for Ub chains F, G, H and I are less well-defined than those for chains E and J, the overall density matches well with individual Ub structures, and thus supports our model. **b**, Two competing molecular replacement (MR) solutions for Ub chain I. During the early stage of molecular replacement, solutions Ib (grey) as well as Ia (red, final solution) received high *Z*-scores, but only Ia was confirmed once the rest of the complex (all four 2CARDs and Ub chains E–H and J) was fixed as a known solution (see Methods). In addition, the simulated annealing omit map from (**a**) matches Ia, not Ib. Based on these results, Ia was selected for our final model. **c**, B-factor representation of the 2CARD–Ub complex. Note that Ub chains F–I have high B-factors, indicative of a high degree of flexibility. See (**d**) for an explanation. **d**, Three types of crystallographic packing interactions. They are mediated by Ub(E)–Ub(J) contacts (left); 2CARD–Ub(E) and 2CARD–2CARD contacts (middle); and 2CARD–Ub(H) contacts (right). Solid lines indicate crystallographic contacts, whereas dotted lines indicate boundaries between adjacent 2CARD tetramers without direct contacts. Note that 2CARD–Ub(H) contacts are less intimate than other contacts. The extensive contacts with Ub(E) and Ub(J) explain lower B-factors observed with these two Ub chains in **c**. **e**, SDS–PAGE analysis of the crystals shows that K63-Ub_2 in the crystal is intact. **f**, Analysis of distance between the C terminus and K63 of adjacent Ub chains. As residues 72–76 of several Ubs were disordered in the structure, we measured the distance between Cα of L71 (black sphere) and Cα of K63 (blue sphere). The distance requirement for covalent conjugation (<25Å) is shown on the right, assuming 3.5Å spacing per residue in the missing C-terminal tail, 2Å for the C-terminal carboxylate, 6Å for Lys side chains. This condition is only satisfied with three pairs of Ubs, which enabled us to unambiguously identify pairs of Ubs that are covalently connected through the K63-linkage. **g**, Two crystallographic packing arrangements, which can potentially allow sharing of a single chain of Ub_2 by two neighbouring 2CARD tetramers. In these arrangements, K63 of Ub bound to one 2CARD tetramer is within the covalent conjugation distance (Cα distance <25Å) from L71 of Ub bound to a neighbouring 2CARD tetramer. Such Ub crossover would increase the heterogeneity in the Ub connectivity, and could have contributed to the poor electron density map corresponding to the K63-linker. The same colour code was used as in **f**.



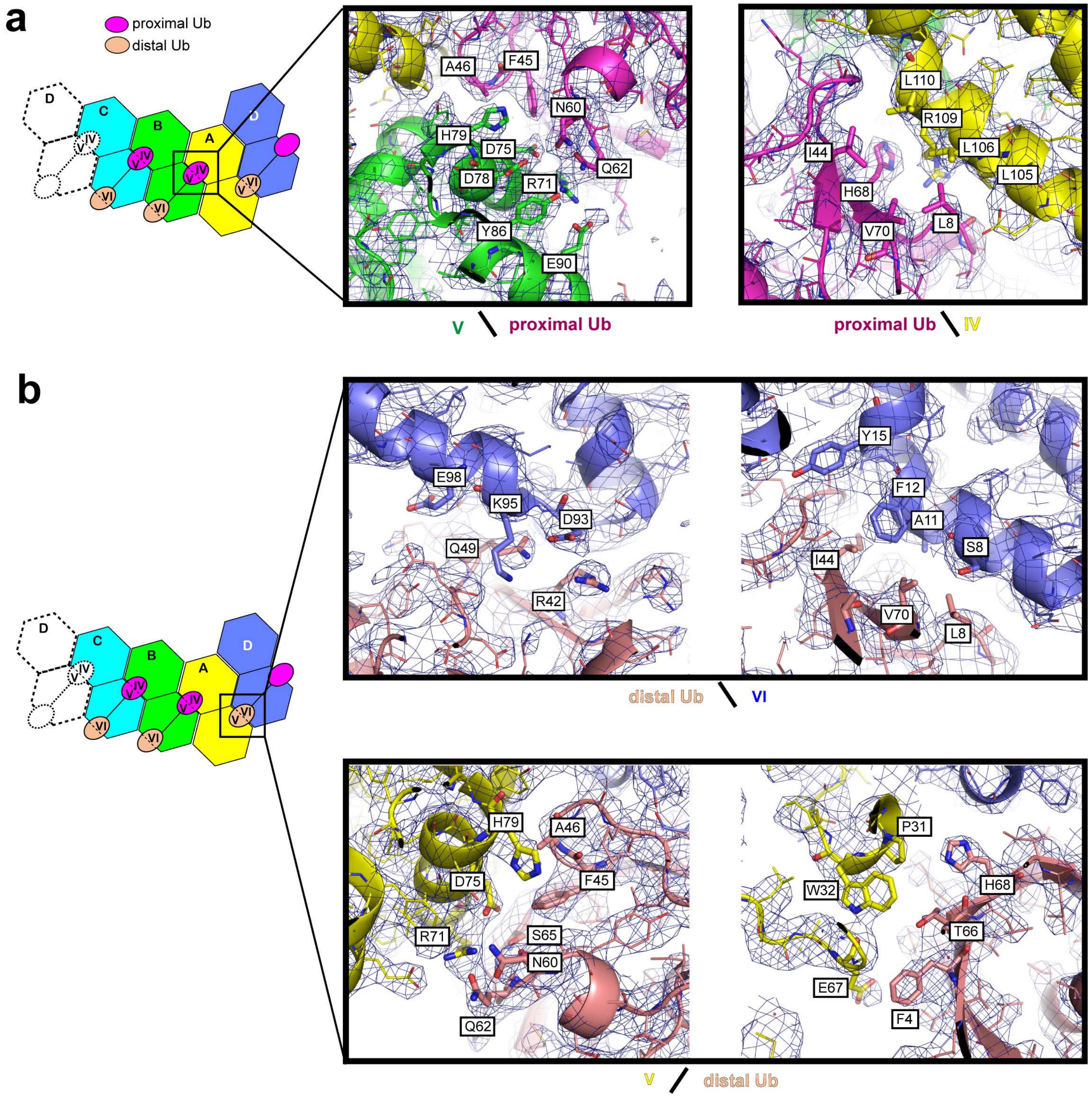
Extended Data Figure 3 | Assembly of the 2CARD tetramer is mediated by rigid-body docking along the helical trajectory with the pitch in the screw equivalent to a single CARD. **a**, Superposition of RIG-I 2CARD in the tetramer (subunits A–D) and 2CARD from full-length duck RIG-I (PDB: 4A2W). **b**, Superposition of full-length RIG-I (grey) onto the 2CARD tetramer by aligning 2CARD in full-length RIG-I with 2CARD subunit A (yellow) in the tetramer. The same colour code was used as in **a**. The superposition shows that the helicase domain in full-length RIG-I masks the 2CARD(A)–2CARD(B) interface and sterically blocks subunit B (green) from interacting with A. On the right, surface representation was separately shown for 2CARD(B) in the tetramer (top) and the helicase domain in full-length RIG-I (bottom) to further demonstrate the steric clash between the helicase domain and 2CARD(B). **c**, Geometric relationship between adjacent 2CARDs. Superposition of the 'cut-out' dimers of A–B, B–C and C–D (left), showing repetition of intermolecular interactions along the helical trajectory. The D–A interaction (middle) in the helical 'seam' (as defined in Fig. 1b) differs from the A–B, B–C and C–D interactions by a relative dislocation of A by a single CARD (right). **d**, IFN-β reporter activity of wild-type RIG-I and tetramer interface mutants (in Fig. 1d) with and without 42 bp dsRNA stimulation in 293T cells (mean ± s.d., $n = 3$).



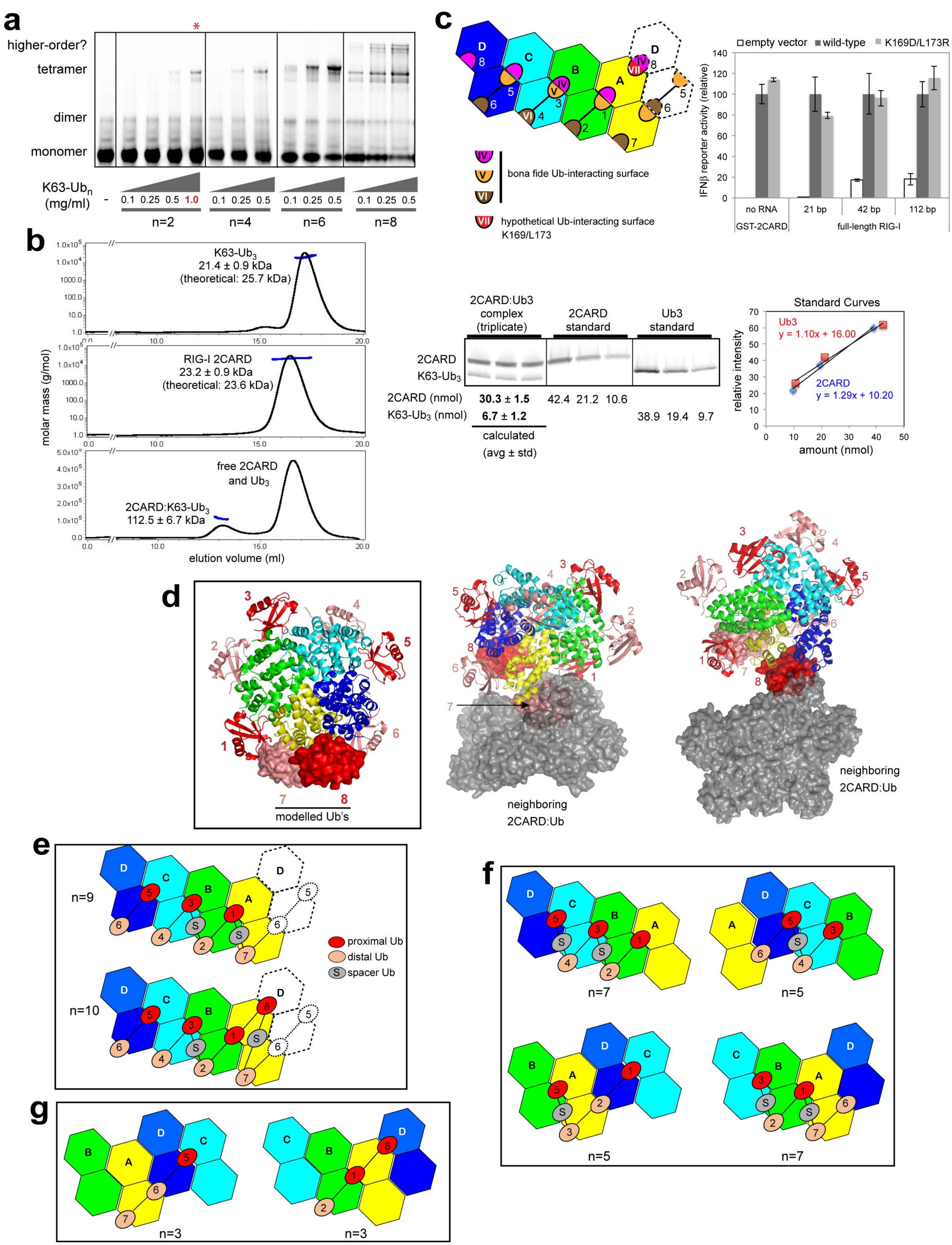
Extended Data Figure 4 | Detailed analysis of the 2CARD tetramerization interface. Electron density map ($2F_o$-F_c) was contoured at $\sigma = 0.9$. A few residues at the interface were displayed as stick models with labels. **a**, Definition of the interaction surface type (same as in Fig. 1c). **b**, Intramolecular interaction between surface IIb and IIa of the first and second CARDs, respectively. **c**, Intermolecular interactions that repeat along the helical trajectory (A–B, B–C and C–D). Shown are the interactions between 2CARD(A) and (B), which are identical to those between (B) and (C), and between (C) and (D) (Extended Data Fig. 3c). These interactions involve three interfaces (shown in the left, middle and right panels). Each of these interfaces consists of a few ($<$ 3–4) residues on each side of the molecules, suggesting that cooperativity of all three interactions might be important for the tetramer stability. The IIIa–IIIb interactions in the first and second CARDs (left and right panels) are more extensive than the Ia–Ib interaction (middle), and are in general electrostatic. Detailed interactions among the interface residues could not be unambiguously determined due to the limited resolution of the structure. **d**, Intermolecular interactions at the helical seam, which occurs only between 2CARD(D) and 2CARD(A). As with the interaction between 2CARD(A) and (B) (or between B and C, and between C and D along the helical trajectory), each of the three interfaces consists of a few ($<$ 3–4) residues on each side of the molecules, and the IIIa–IIIb interaction is more extensive than either of the two 1a–Ib interactions. Note that IIIa of the second CARD interacts with IIIb of the first CARD at the helical seam, whereas it interacts with IIIb of the second CARD along the helical trajectory (**c**). This is despite the low level of conservation of IIIb between the first and second CARDs, and thus suggests plasticity in molecular interactions. Similarly, Ia of the second CARD also has two distinct interaction partners, that is, Ib of the second CARD at the helical seam and Ib of the first CARD along the helical trajectory (**c**). But in this case, Ia utilizes different residues (albeit in the same local area) to accommodate different interaction partners.



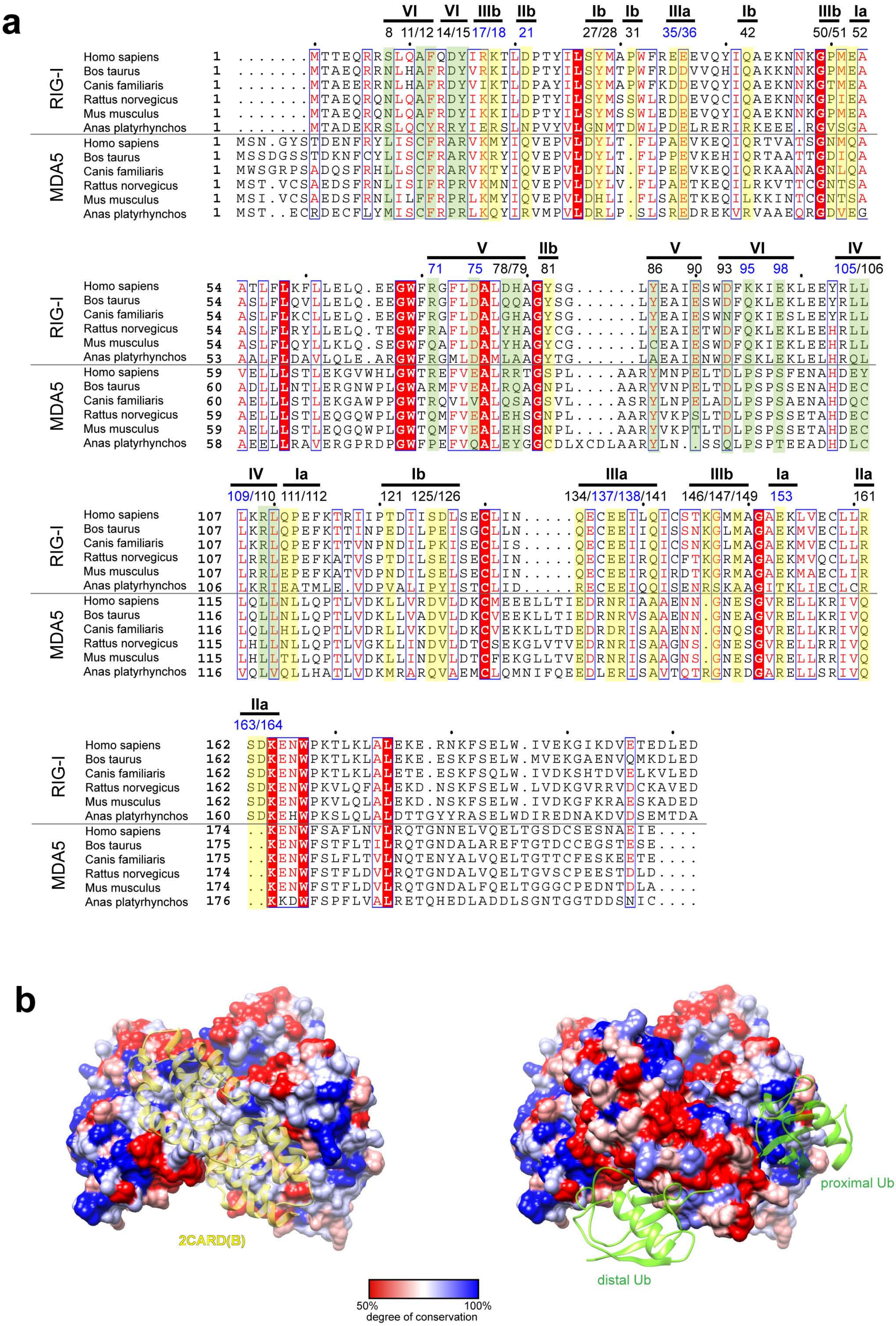
Extended Data Figure 5 | Interaction between the 2CARD tetramer and K63-Ub$_2$. **a**, Superposition of RIG-I 2CARD in the tetramer (subunits A–D) bound with K63-Ub$_2$. The three Ub$_2$ chains bound to A–B, B–C and C–D–A were superposed by aligning chain A, B and C, respectively. **b**, Superposition of the first (dark grey) and second (light grey) CARDs from RIG-I 2CARD bound by proximal (red) or distal (salmon) Ubs. The good alignment suggests that the proximal Ub-binding site, IV, in the second CARD is equivalent to the distal Ub-binding site, VI, in the first CARD (see Fig. 3b for the definitions of IV and VI). **c**, Distance analysis between the C-terminal residue of the distal Ub and K63, K48 and M1 of the proximal Ub. The C-terminal tail (residues 74–76) of the distal Ub is disordered in the structure, and thus the distance was measured between Cα of residue 73 and Cα of the target Lys. Below is the distance requirement for covalent conjugation, assuming 3.5Å spacing per residue in the missing C-terminal tail, 2Å for the C-terminal carboxylate, 6Å for Lys side chains and 1.5Å for N-terminal amine. These distance requirements are satisfied only with K63, thus rationalizing the observed specificity of 2CARD for the K63-linkage. **d**, IFN-β reporter activity of full-length RIG-I stimulated with 21 bp, 42 bp and 112 bp dsRNAs or isolated 2CARD fused to GST (GST–2CARD) without RNA. Activities were compared between wild type and mutants defective in Ub binding (R71D/D75R and K95E/E98R) or conjugation (6KR), and were normalized against the wild-type values (mean ± s.d., $n = 3$). 6KR indicates Arg mutation of six Lys (K99, K169, K172, K181, K190 and K193) that are known to be conjugated with K63-Ub$_n$. The signalling activity of full-length RIG-I with 21 bp dsRNA was completely abrogated by R71D/D75R, K95E/E98R or 6KR, suggesting the importance of both Ub-binding and conjugation. The negative impacts of K95E/E98R and 6KR were progressively alleviated by stimulation with increasing length of dsRNA (42 and 112 bp). This restoration of the signalling activity by longer dsRNAs is consistent with our previous report that filament formation of RIG-I on long dsRNA (>60 bp) promotes 2CARD tetramerization by the 'proximity-induced' mechanism[12]. Note that 21 bp, 42 bp and 112 bp dsRNA can accommodate 1–2, 3, 8 RIG-I molecules, respectively. The negative impact of R71D/D75R could not be alleviated by stimulation with longer dsRNAs, which is somewhat at odds with the result with another Ub-binding deficient mutant, K95E/E98R. It is possible that R71D/D75R has more severe defects (in Ub binding or perhaps in 2CARD structure), which could not be overcome by Ub-conjugation or filament formation. For comparison, we also used GST–2CARD, which has been widely used in previous studies[1–3]. Despite the fact that GST–2CARD cannot form filaments, its sensitivity to the mutations was equivalent to full-length RIG-I with 42 bp, rather than 21 bp dsRNA. This likely reflects the effect of the fusion partner, GST, which forms a constitutive dimer. Isolated 2CARD without GST is a very poor stimulant of IFN-β, thus could not be used in this study.



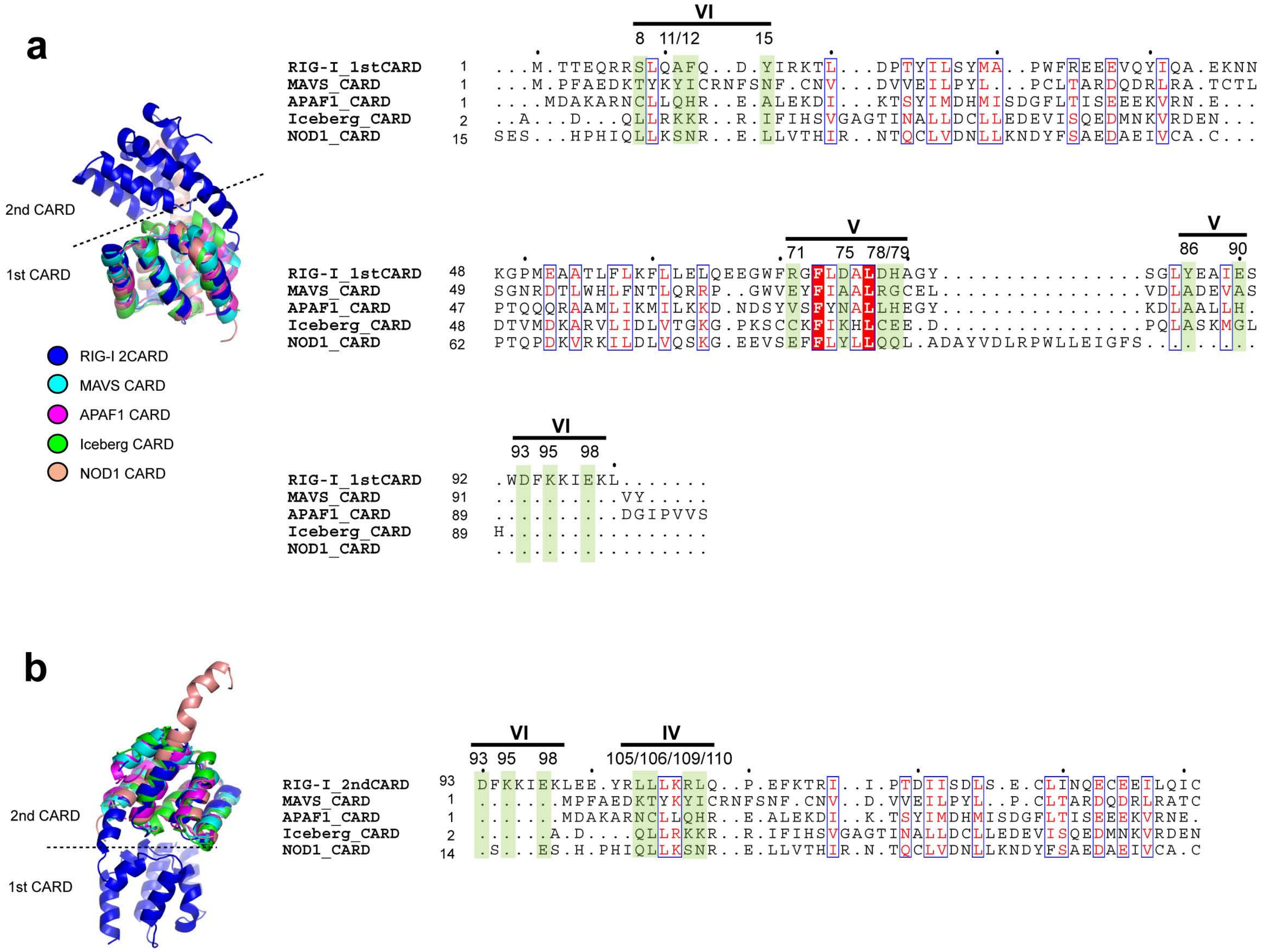
Extended Data Figure 6 | Detailed analysis of the 2CARD–Ub interface. Electron density map ($2F_o$-F_c) was contoured at $\sigma = 0.9$. Surface types were defined in the 2D representation on the left (as in Fig. 3b). **a**, Proximal Ub occupies surfaces IV and V on 2CARD. Surface IV interacts with the hydrophobic patch (L8/I44/V70) of proximal Ub, whereas V interacts with a combination of hydrophilic (N60/Q62) and hydrophobic (F45/A46) residues of proximal Ub. **b**, The distal Ub bound to 2CARD(D) simultaneously occupies both surfaces VI and V, whereas distal Ub's bound to 2CARD(B) and (C) occupy surface VI alone (due to the lack of adjacent V). The interaction between distal Ub and VI is identical no matter whether the same Ub forms an additional interaction with V. Thus, only the distal Ub–2CARD(D) interaction is shown. Surface VI forms a combination of hydrophilic and hydrophobic interactions with Q49/R42 and L8/I44/V70 on distal Ub, respectively. Surface V interacts with distal Ub in the same manner as with proximal Ub, forming contacts with F45/A46/N60/N62 of Ub. Additional interactions involving surface V were seen with F4/T66/H68 of both distal and proximal Ubs (not shown in **a**), but they appear less intimate than those involving F45/A46/N60/N62.


a
higher-order?
tetramer
dimer
monomer
K63-Ubn (mg/ml)
-
0.1 0.25 0.5 1.0
0.1 0.25 0.5
0.1 0.25 0.5
0.1 0.25 0.5
n=2
n=4
n=6
n=8
b
molar mass (g/mol)
K63-Ub3
21.4 ± 0.9 kDa
(theoretical: 25.7 kDa)
RIG-I 2CARD
23.2 ± 0.9 kDa
(theoretical: 23.6 kDa)
free 2CARD
and Ub3
2CARD:K63-Ub3
112.5 ± 6.7 kDa
elution volume (ml)
c
bona fide Ub-interacting surface
hypothetical Ub-interacting surface
K169/L173
empty vector
wild-type
K169D/L173R
IFNβ reporter activity (relative)
no RNA
21 bp
42 bp
112 bp
GST-2CARD
full-length RIG-I
2CARD:Ub3
complex
(triplicate)
2CARD
standard
Ub3
standard
2CARD
K63-Ub3
2CARD (nmol)
30.3 ± 1.5
42.4 21.2 10.6
K63-Ub3 (nmol)
6.7 ± 1.2
38.9 19.4 9.7
calculated
(avg ± std)
Standard Curves
Ub3
y = 1.10x + 16.00
2CARD
y = 1.29x + 10.20
relative intensity
amount (nmol)
d
modelled Ub's
neighboring
2CARD:Ub
neighboring
2CARD:Ub
e
n=9
n=10
proximal Ub
distal Ub
spacer Ub
f
n=7
n=5
n=5
n=7
g
n=3
n=3

Extended Data Figure 7 | High avidity interaction is required for efficient formation of the 2CARD tetramer by K63-Ub$_n$ ($n > 2$). **a**, EMSA analysis of the 2CARD tetramer formation using fluorescently labelled 2CARD-S (50 μM) with increasing concentrations of K63-Ub$_n$ ($n = 2, 4, 6, 8$). Note that an additional higher concentration (1 mg ml^{-1}) was included only for K63-Ub$_2$ (red asterisk), due to its low efficiency to stimulate 2CARD tetramerization. With K63-Ub$_8$, additional bands appeared above the tetramer band, possibly reflecting two or more 2CARD tetramers bridged by a single Ub chains. **b**, Molecular mass analysis of 2CARD in complex with K63-Ub$_3$ using multi-angle light scattering (MALS) coupled to size exclusion chromatography (SEC). Molecular mass estimated for the complex is 112.5 kDa ($\pm$ 6.7 kDa), which is consistent with a tetrameric 2CARD (92.8 kDa, 23.2 $\pm$ 0.9 kDa as a monomer) bound by a single chain of K63-Ub$_3$ (21.4 $\pm$ kDa). This 4:1 binding ratio of 2CARD to Ub$_3$ is further supported by the SDS–PAGE intensity analysis of the complex (purified from MALS-SEC above) using Krypton fluorescence stain (right) (mean $\pm$ s.d., $n = 3$), which suggests the molar ratio of 4.5:1 for 2CARD–Ub$_3$. This result suggests the sufficiency of a single chain of K63-Ub$_n$ ($n >= 3$) for stabilizing the 2CARD tetramer, although it does not exclude potential binding of additional Ubs at saturating concentrations[3]. Note that previous study[3] suggesting 4:4 binding of 2CARD to Ub$_n$ ($n = 3$–6) was performed in a buffer lacking salt (20 mM Tris-HCl (pH 7.5) and 1 mM DTT), whereas the current study was performed with 150 mM NaCl (20 mM HEPES (pH 7.5) and 150 mM NaCl), which could be responsible for the divergent results. The sufficiency of the single chain of Ub$_3$ for stabilizing the 2CARD tetramer suggests that multiple stoichiometries and Ub$_n$ binding configurations are possible, depending on the concentrations of 2CARD and Ub$_n$ as well as buffer compositions. See **g** for how a single chain of K63-Ub$_3$ could stabilize the 2CARD tetramer. **c**, Six Ub binding sites (1–6) were occupied in the crystal structure, with a potential for binding of up to eight total Ub molecules in the 2CARD tetramer. Site 7 is equivalent to sites 2, 4 and 6, and thus is a bona fide Ub-binding site. Site 8 is a hypothetical Ub-binding site, as its interaction with Ub would simultaneously utilize surfaces IV and VII, instead of surfaces IV and V as in sites 1, 3 and 5. Mutations of VII (K169D/L173R) did not affect the signalling activity of RIG-I in cells based on the IFNβ reporter assay (right) (mean $\pm$ s.d., $n = 3$). Although this result suggests that surface VII (which only affects site 8, not 1–7) is not important, it does not exclude the possibility of site 8 serving as another Ub-binding site, as the loss of one out of 8 sites may not have a significant effect on the signalling outcome. **d**, A model of the 2CARD tetramer with all 8 potential Ub binding sites occupied. Ub bound to sites 7 and 8 (surface representation) were modelled by superposing 2CARDs bound to distal and proximal Ub onto 2CARD(A) and (D), respectively. Ub binding sites are numbered according to the 2D representation in **c**. Crystallographic packing prevents Ub occupancy of sites 7 and 8. Neighbouring molecules, which occlude the sites 7 (left) and 8 (right), are shown in grey surface. **e**, Two configurations to occupy 7 or 8 potential Ub-binding sites in the 2CARD tetramer using a single K63-Ub$_n$ chain. Ub-binding sites are numbered as in **b**. Ub labelled 'S' stands for the unbound Ub that serves as a spacer. The presence of spacer Ub is consistent with the observed activity of Ub$_4$ with K63- and K48- mixed linkage in stimulating RIG-I 2CARD[3], as there is no geometric restriction to impose the linkage specificity for the spacer Ub. **f**, Minimal length of K63-Ub$_n$ chain that allows bridging of four 2CARDs by a single Ub chains. Four examples were shown, in which Ub chains start with 2CARD subunit A, B, C or D. Ub-binding sites are numbered as in **b**. Ub labelled 'S' stands for the unbound Ub that serves as a spacer. **g**, Two configurations that a single chain of K63-Ub$_3$ can bridge three 2CARDs (without involvement of a spacer Ub).


a
RIG-I
MDA5
Homo sapiens
Bos taurus
Canis familiaris
Rattus norvegicus
Mus musculus
Anas platyrhynchos
b
2CARD(B)
proximal Ub
distal Ub
50%
100%
degree of conservation

Extended Data Figure 8 | Sequence conservation analysis of the 2CARD–2CARD and 2CARD–Ub interface. **a**, Sequence alignment of RIG-I and MDA5 2CARD (using the program ClustalOmega[27]). Residues in RIG-I involved in the 2CARD–2CARD interactions (surface Ia/b-IIIa/b in Extended Data Fig. 4) and Ub binding (surface IV, V and VI in Extended Data Fig. 6), and their equivalent residues in MDA5 are highlighted (yellow and green for 2CARD–2CARD and 2CARD–Ub interfaces, respectively). Residue numbers right above highlights are according to human RIG-I. Residues tested by mutagenesis in this study are indicated by blue colour. Residues involved in the 2CARD–2CARD interactions show, on average, a moderate level of conservation in comparison to other residues on the surface of RIG-I 2CARD (see surface representation in **b**). Poor conservation of the tetramerization interface is consistent with previous observations that protein–protein interfaces often display evolutionary versatility due to the plasticity of the interaction and co-evolution of the interacting surfaces[28,29]. The Ub binding surface of RIG-I is more conserved than the 2CARD–2CARD interface (see surface representation in **b**), possibly reflecting the conserved nature of Ub. When the comparison is made between RIG-I and MDA5, we found that only four residues are well-conserved (F12 and L110 in the Ub-binding surface, and E36 and K164 in tetramerization surface), assuming interchange of residues within each group of F/Y, L/V/I or D/E as well-conserved. This is insufficient to support structural conservation of the 2CARD tetramer or 2CARD–Ub complex between RIG-I and MDA5. The structure of the MDA5 oligomers and/or MDA5–Ub complex would be required to compare 2CARD oligomerization mechanism between RIG-I and MDA5. **b**, Degree of sequence conservation (within RIG-I based on **a**) mapped onto the RIG-I 2CARD tetramer structure (generated using the program Chimera). Consistent with the analysis above, the 2CARD–2CARD interfaces show a moderate level of conservation, whereas Ub-binding sites show a higher degree of conservation. Other conserved surface areas may be involved in interactions with other molecules, such as TRIM25 or MAVS.



Extended Data Figure 9 | Sequence analysis of the Ub-binding surface in the CARD family. Structure-based sequence alignment (using the program SALGIN[30]) of various CARD domains. In our effort to further analyse potential generality of 2CARD–Ub interaction observed with RIG-I in our structure, we aligned other CARDs with RIG-I 2CARD. We performed structure-based sequence alignment, as many members of the CARD family share little sequence similarity. Three dimensional protein structure, which is more conserved than the primary sequence, allows more accurate sequence comparison. The first CARD (**a**) and second CARD (**b**) of human RIG-I was aligned with other CARDs from MAVS, APAF1, Iceberg and NOD1 (PDB code: 2VGQ, 3YGS, 1DGN and 4E9M, respectively). None of the Ub-binding residues in RIG-I 2CARD are conserved in these CARDs.

Extended Data Table 1 | Data collection and refinement statistics

	Crystal 1
Data collection	
Space group	P1 2_1 1
Cell dimensions	
a, *b*, *c* (Å)	83.5, 101.9, 88.2
α, β, γ (°)	90.0, 106.9, 90.0
Resolution (Å)	50-3.7 (3.93-3.70)*
R_{sym} or R_{merge}	0.165 (0.621)
$I / \sigma I$	5.92 (1.55)
Completeness (%)	98.1 (96.7)
Redundancy	4.0 (4.0)
Refinement	
Resolution (Å)	46.12-3.70 (3.82-3.70)
No. reflections	29885 (2697)
R_{work} / R_{free}	0.222 (0.333) / 0.285 (0.401)
No. atoms	
Protein	19179
Ligand/ion	
Water	
B-factors	
Protein	92.3
Ligand/ion	
Water	
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.6

* Values in parentheses are for highest-resolution shell.

doi:10.1038/nature13083

Structure of the human P2Y$_{12}$ receptor in complex with an antithrombotic drug

Kaihua Zhang[1]*, Jin Zhang[1]*, Zhan-Guo Gao[2], Dandan Zhang[1], Lan Zhu[1], Gye Won Han[3], Steven M. Moss[2], Silvia Paoletta[2], Evgeny Kiselev[2], Weizhen Lu[1], Gustavo Fenalti[3], Wenru Zhang[1], Christa E. Müller[4], Huaiyu Yang[5], Hualiang Jiang[5], Vadim Cherezov[3], Vsevolod Katritch[3], Kenneth A. Jacobson[2], Raymond C. Stevens[3,6], Beili Wu[1] & Qiang Zhao[1]

P2Y receptors (P2YRs), a family of purinergic G-protein-coupled receptors (GPCRs), are activated by extracellular nucleotides. There are a total of eight distinct functional P2YRs expressed in human, which are subdivided into P2Y$_1$-like receptors and P2Y$_{12}$-like receptors[1]. Their ligands are generally charged molecules with relatively low bioavailability and stability *in vivo*[2], which limits our understanding of this receptor family. P2Y$_{12}$R regulates platelet activation and thrombus formation[3,4], and several antithrombotic drugs targeting P2Y$_{12}$R—including the prodrugs clopidogrel (Plavix) and prasugrel (Effient) that are metabolized and bind covalently, and the nucleoside analogue ticagrelor (Brilinta) that acts directly on the receptor—have been approved for the prevention of stroke and myocardial infarction. However, limitations of these drugs (for example, a very long half-life of clopidogrel action and a characteristic adverse effect profile of ticagrelor)[5,6] suggest that there is an unfulfilled medical need for developing a new generation of P2Y$_{12}$R inhibitors[7,8]. Here we report the 2.6 Å resolution crystal structure of human P2Y$_{12}$R in complex with a non-nucleotide reversible antagonist, AZD1283. The structure reveals a distinct straight conformation of helix V, which sets P2Y$_{12}$R apart from all other known class A GPCR structures. With AZD1283 bound, the highly conserved disulphide bridge in GPCRs between helix III and extracellular loop 2 is not observed and appears to be dynamic. Along with the details of the AZD1283-binding site, analysis of the extracellular interface reveals an adjacent ligand-binding region and suggests that both pockets could be required for dinucleotide binding. The structure provides essential insights for the development of improved P2Y$_{12}$R ligands and allosteric modulators as drug candidates.

P2Y$_{12}$R, a member of the P2Y purinergic GPCR family stimulated by adenosine diphosphate (ADP), is a major player in platelet aggregation and granule secretion and supports the formation of a thrombus[9]. Ethyl 6-(4-((benzylsulphonyl)carbamoyl)piperidin-1-yl)-5-cyano-2-methylnicotinate (AZD1283) (Extended Data Fig. 1) was recently revealed by AstraZeneca as a novel P2Y$_{12}$R antagonist for the treatment of arterial thrombosis. AZD1283 efficiently inhibits platelet activation *in vivo* with only a limited increase in the bleeding time[10].

To understand how antithrombotic drugs recognize their purinoceptor target, we solved the structure of an engineered human P2Y$_{12}$R by inserting a thermostabilized apocytochrome, b_{562} RIL (BRIL), in the third intracellular loop (ICL3)[11] in complex with AZD1283 at 2.6 Å (Extended Data Table 1). A point mutation—D294$^{7.49}$N (superscript indicates residue numbering using Ballesteros–Weinstein nomenclature[12])—in the N[D]P$^{7.50}$xxY motif of helix VII was introduced to improve the purified protein yield twofold (Extended Data Fig. 2). Saturation and competition binding assays indicate that neither the BRIL fusion nor the point mutation significantly influenced ligand binding of P2Y$_{12}$R (Extended Data Table 2).

The overall fold of the P2Y$_{12}$R structure consists of a canonical seven transmembrane (7TM) bundle of α-helices and a carboxy-terminal helix VIII that is parallel to the membrane bilayer (Fig. 1a). Several loops, especially ECL2, appear to be flexible and result in a total of 24

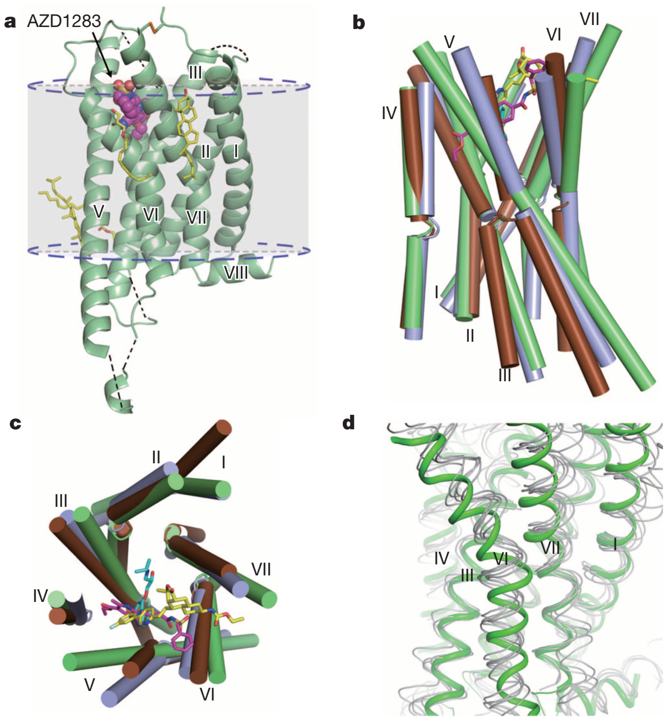


Figure 1 | Overview of the P2Y$_{12}$R–AZD1283 complex structure. a, Cartoon representation of P2Y$_{12}$R. P2Y$_{12}$R is coloured green. AZD1283 is shown as magenta spheres. Cholesterol and lipids have yellow carbons. The disulphide bridge is shown as lime sticks. Missing loops and membrane boundaries are indicated as black and blue dashed lines, respectively. **b**, **c**, Side (**b**) and top (**c**) views of P2Y$_{12}$R (green cylinders) compared with β$_2$AR (PDB accession 2RH1, brown) and PAR1 (PDB accession 3VW7, blue). The ligands AZD1283, carazolol and vorapaxar are shown as sticks with magenta, cyan and yellow carbons, respectively. Other elements are coloured as follows: oxygen, red; nitrogen, dark blue; sulphur, yellow. **d**, Comparison of P2Y$_{12}$R with β$_2$AR, PAR1, A$_{2A}$ adenosine receptor (A$_{2A}$AR; PDB accession 3EML), neurotensin receptor (NTSR1; PDB accession 4GRV) and κ opioid receptor (κ-OR; PDB accession 4DJH). P2Y$_{12}$R is shown in green and the other GPCRs are in grey.

[1]CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Pudong, Shanghai 201203, China. [2]Molecular Recognition Section, Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [4]PharmaCenter Bonn, Pharmaceutical Institute, Pharmaceutical Chemistry I, An der Immenburg 4, D-53121 Bonn, Germany. [5]Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Pudong, Shanghai 201203, China. [6]iHuman Institute, ShanghaiTech University, 99 Haike Road, Pudong, Shanghai 201203, China.
*These authors contributed equally to this work.

unmodelled loop residues (88–91 in ECL1, 133–135 intracellular loop 2 (ICL2), 163–178 ECL2 and 230 in ICL3). Only one disulphide bond is clearly observed in the structure, connecting the amino terminus (C17) with helix VII ($C270^{7.25}$). Two receptor molecules from adjacent unit cells form receptor–receptor interactions in a parallel orientation mediated by helix V. Two cholesterol molecules are observed bound to each receptor: one is at the interface of helices III and V, stabilizing the receptor–receptor interaction, and the other is at the interface of helices I and VII and does not participate in crystal contacts (Extended Data Fig. 3).

The $P2Y_{12}R$ structure has important features that set it apart from another representative of the δ group of class A GPCRs[13], protease activated receptor 1 (PAR1, ~24% sequence identity; transmembrane C_α root mean squared deviation (r.m.s.d.) ~2.2 Å), and other known class A structures, such as the β_2-adrenergic receptor (β_2AR, transmembrane C_α r.m.s.d. 2.6–3.5 Å) (Fig. 1b–d). Whereas helix V in most of the class A GPCR structures that have been determined so far is bulged and bent at a highly conserved $P^{5.50}$, $P2Y_{12}R$ has $N201^{5.50}$ (and $V^{5.50}$ for the other $P2Y_{12}R$-like receptors, $P2Y_{13}R$ and $P2Y_{14}R$), and, consequently, its structure lacks the corresponding helical bend. Moreover, $P2Y_{12}R$ has no other proline or glycine residues in helix V that could destabilize its straight α-helical conformation. Besides $P2Y_{12}R$, the only other receptor structure solved so far that also lacks the conserved proline and corresponding bend in helix V is the sphingosine 1-phosphate receptor 1 ($S1P_1$)[14]. However, compared with $S1P_1$, helix V of $P2Y_{12}R$ extends ~2 additional helical turns above the extracellular side of the membrane, and thus is also ~1–2 helical turns longer than in other known GPCR structures. The unique straightening and elongation of helix V observed in the $P2Y_{12}R$ structure results in a shift of its extracellular end towards helix IV by more than 6 Å in comparison with other known class A structures, leading to rearrangements in other helices and extracellular loops.

Other distinct conformational features of the $P2Y_{12}R$ structure are observed in helices VII and VI. The intracellular tip of helix VII is closer to the axis of the 7TM bundle than in most class A GPCR structures, in an 'inward' position similar to PAR1 (ref. 15). The helix VI intracellular tip is shifted slightly outward, and the whole helix VI is translated along its axis towards the intracellular surface by a half α-helical turn as compared with other known GPCR structures (Fig. 1d and Extended Data Fig. 4d). Owing to this shift, the conserved $R122^{3.50}$ in the $D(E)R^{3.50}Y$ motif, which forms an ionic or a hydrogen-bonding lock with $E^{6.30}$ or $T^{6.34}$ in rhodopsin and several other GPCRs, is actually located at the same level as a hydrophobic side-chain $V238^{6.37}$ in the $P2Y_{12}R$ structure (Extended Data Fig. 4a–c). This residue arrangement excludes formation of an ionic lock or polar interaction between helix VI and the DRY motif. The lack of an ionic lock or a polar interaction between helix VI and the DRY motif, as well as active-like conformations of the intracellular side of helices VI and VII, suggest that $P2Y_{12}R$ might be more prone to activation, in agreement with previous reports that this receptor exhibits high levels of basal activity[16,17]. Another key structural difference in helix VI occurs in the centre of the 7TM helical bundle at position 6.48, which in many class A GPCRs contains a tryptophan residue, involved in ligand binding and implicated in signal transduction[18]. In all P2YRs and PARs, however, $W^{6.48}$ is replaced with $F(Y)^{6.48}$, which displays the same orientation in both structures that might be a common feature within the δ group (Extended Data Fig. 4e).

AZD1283 binds to $P2Y_{12}R$ in a pocket that is very distinct in location and shape from that seen in previously solved GPCRs (Fig. 2). The elongated ligand stretches over more than 17 Å between helices IV and VII, making a number of polar and hydrophobic interactions with side chains from helices III–VII (Fig. 2a, b). The orientation of this ligand is orthogonal to a general ligand position in solved receptor complexes from α, β and γ groups in class A GPCRs, although it shares some similarity with the orientation of vorapaxar in PAR1, which belongs to the δ group (Extended Data Fig. 4). The piperidinyl–nicotinate group of the antagonist inserts into a sub-pocket formed by helices III, IV and V, whereas the benzylsulphonyl group mainly interacts with helices VI and VII. The structure reveals at least seven polar and ionic interactions between $P2Y_{12}R$ and AZD1283 (Fig. 2b). Additionally, the benzene ring of $Y105^{3.33}$ forms a π–π interaction with the nicotinate and a hydrophobic interaction with the piperidine group, and the phenyl group of the ligand inserts into a hydrophobic pocket formed by the side chains of $F252^{6.51}$, $R256^{6.55}$, $Y259^{6.58}$, $L276^{7.31}$ and $K280^{7.35}$.

Although the AZD1283-binding pocket of $P2Y_{12}R$ and its equivalent in PAR1 are formed by the same helices, there are significant differences between them. Whereas in PAR1, 24 residues of ECL2 cover the entire

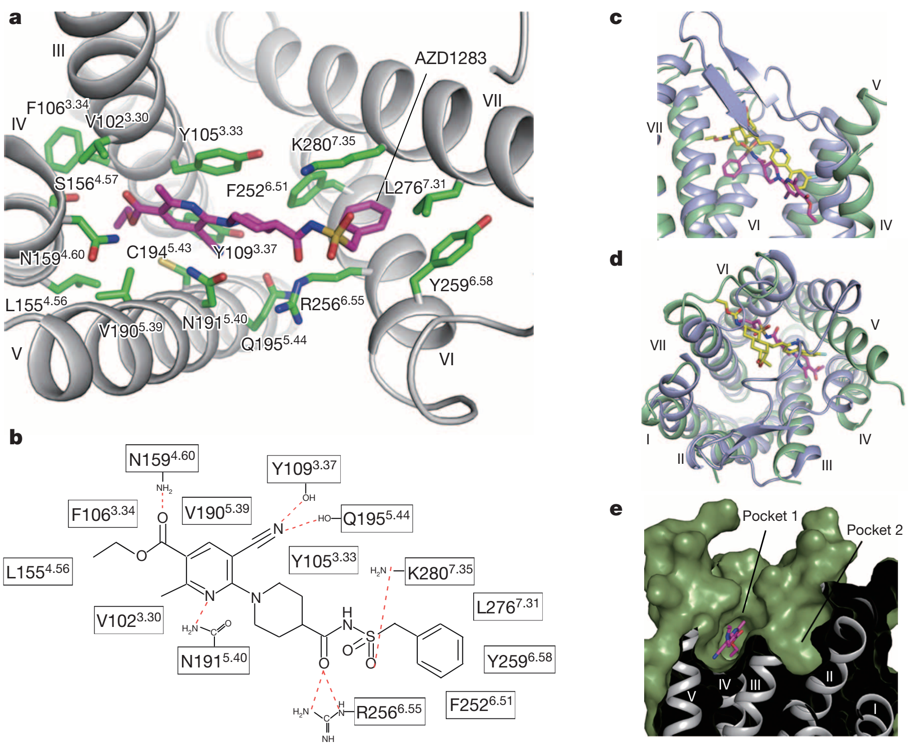


Figure 2 | $P2Y_{12}R$ ligand-binding pocket for AZD1283. a, Key residues in $P2Y_{12}R$ for AZD1283 binding. AZD1283 (magenta carbons) and receptor residues (green carbons) involved in ligand binding are shown in stick representation. **b**, Schematic representation of interactions between $P2Y_{12}R$ and AZD1283. Polar interactions are shown as red dashed lines. **c**, **d**, Side view (**c**) and top view (**d**) comparison of ligand-binding sites in $P2Y_{12}R$ (green) and PAR1 (blue). The antagonists AZD1283 and vorapaxar have magenta and yellow carbons, respectively. **e**, Side view of the $P2Y_{12}R$ ligand-binding pocket. The receptor is shown in both green surface and grey ribbons, and AZD1283 is shown in sticks. Pocket 1 and pocket 2 are indicated by black lines.

binding pocket and participate in extensive ligand interactions, the unresolved 16 residues of ECL2 in P2Y$_{12}$R are likely to be more flexible and to have a lesser role in interactions with AZD1283. Moreover, the extracellular ends of helices IV, VI and VII in P2Y$_{12}$R are shifted outwards compared with PAR1, making the ligand-binding pocket of P2Y$_{12}$R more open and allowing AZD1283 to bind deeper in the 7TM domain (Fig. 2c, d). Within the P2Y$_{12}$R extracellular cavity formed by all 7TM helices, two residues, Y105$^{3.33}$ and K280$^{7.35}$, form a barrier separating the cavity into two pockets. Pocket 1 is composed of helices III–VII, forming the binding site of AZD1283, whereas pocket 2 consists of helices I–III and VII and is not occupied in the structure (Fig. 2e).

Two cysteine residues in helix III and ECL2 (C97$^{3.25}$ and C175 in P2Y$_{12}$R) are highly conserved throughout the GPCR superfamily[19], and, when present, they are observed to form a disulphide bond in all the GPCR structures solved so far[20,21]. To our knowledge, the only receptor structure that lacks the conserved Cys$^{3.25}$ and the corresponding disulphide bridge is S1P$_1$ (ref. 14), and it also lacks any secondary structure in ECL2. In the crystal structure of P2Y$_{12}$R, no electron density is observed for most of ECL2 or for a potential disulphide bond at C97, suggesting the possibility of a labile or dynamic disulphide bond in P2Y$_{12}$R, a feature that would be different from other known GPCRs and may be important functionally. Consistent with the labile nature of a disulphide in the structure, C97$^{3.25}$A and C175A mutations retain similar protein yield and stability, with no obvious aggregation, according to analytical size-exclusion chromatography (aSEC) (Extended Data Fig. 5c). In addition, the melting temperatures of both mutants are substantially higher in complex with AZD1283 than in the apo form and are equivalent to that of the native constructs (Extended Data Fig. 5f), suggesting that neither the receptor's conformation nor its ability to bind AZD1283 is affected by the cysteine mutations. On the contrary, alanine mutations of C17 or C270$^{7.25}$, which disrupt the disulphide bond that is clearly observed in the structure, lead to a marked receptor aggregation (Extended Data Fig. 5d). These data suggest that the disulphide bridge between ECL2 and helix III may be labile in native P2Y$_{12}$R, thus leaving C97$^{3.25}$ and C175 available for interactions with the thiol moieties of drug metabolites.

Residue C97$^{3.25}$ has been previously implicated by functional assays as a covalent binding site for the active metabolites of P2Y$_{12}$R drugs[8,22]. Our aSEC results are consistent with C97$^{3.25}$ being the primary attachment site for the active metabolite of prasugrel, R-138727 (Extended Data Fig. 5e). The suggestion of a dynamic disulphide bridge in our structure correlates with the previous observation that thiol-reactive reagents such as *p*-chloromercuribenzene sulphonate (pCMBS) specifically inhibit activation of P2Y$_{12}$R but not other receptors such as P2Y$_1$R[23]. In addition, these observations are consistent with the absolute selectivity of the active metabolites of prasugrel and clopidogrel for the P2Y$_{12}$R; so far, there have been no reports that they exhibit detectable binding to any other receptor. The structure of P2Y$_{12}$R opens up more possibilities to explain these previous observations, but additional evidence is needed to fully understand their mechanisms.

As residue C97$^{3.25}$ belongs to pocket 2 in the AZD1283-bound P2Y$_{12}$R, the active metabolites of drugs that covalently link to this residue may occupy pocket 2 as well. This is consistent with the docking analysis, from which pocket 2 seems to be an energetically favourable binding site for these active metabolites (Fig. 3a, b). However, docking of nucleotide agonists into the crystal structure of antagonist-bound P2Y$_{12}$ is less definitive, which probably reflects high conformational plasticity of the ligand-binding region in this receptor. Although limited mutagenesis data are consistent with binding of the nucleotides in pocket 2, docking allows for 2-methylthio-adenosine 5′-diphosphate (2MeSADP) binding in either pocket (Extended Data Fig. 6). Mutations of a few residues that belong to pocket 2, K80$^{2.60}$A or E281$^{7.36}$A, or a residue at the interface of the two pockets, K280$^{7.35}$A, decrease the binding affinity of the nucleotide radioligand [^{3}H]2MeSADP (Extended Data Table 3 and Extended Data Fig. 7). In both docking models, R256$^{6.55}$, which was previously reported to be important for the activation of P2Y$_{12}$R[24,25], potentially makes contact with the phosphate groups of 2MeSADP. This is consistent with pharmacological and biochemical data showing that mutations of this residue affect both the number of binding sites and their affinities for radio-labelled 2MeSADP in patient platelets and in transfected CHO cells[26–28]. Another residue, R265^{ECL3}, which was previously reported to affect P2Y$_{12}$R activation, is located on ECL3 and should not directly interact with

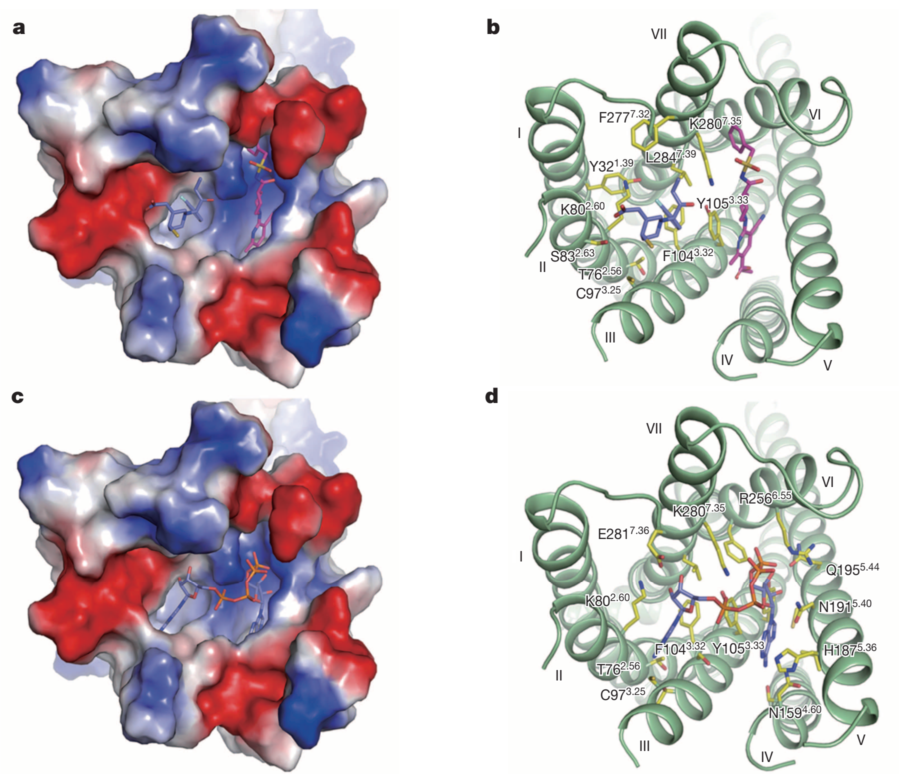


Figure 3 | Hypothetical binding modes of other ligands to P2Y$_{12}$R obtained by docking simulations. **a**, **b**, Electrostatic surface representation (**a**) and cartoon representation (**b**) of the docking model of R-138727, with (*R*)-configuration of the reactive thiol-group and (*S*)-configuration at the benzylic position[29], binding to P2Y$_{12}$R (before covalent bond formation). The surface is coloured according to its electrostatic potential from red (negative) to blue (positive). The receptor is shown in green and R-138727 is represented as sticks with blue carbons. Side chains of residues involved in the binding of R-138727 are shown as sticks (yellow carbons). AZD1283 is shown as sticks with magenta carbons. **c**, **d**, Electrostatic surface representation (**c**) and cartoon representation (**d**) of the docking model of diadenosine tetraphosphate (Ap4A) binding to P2Y$_{12}$R. Ap4A is represented as sticks with blue carbons, and the receptor and residues involved in binding are coloured as in **b**). Other elements are coloured as follows: red, oxygen; dark blue, nitrogen; yellow, sulphur; orange, phosphorus.

2MeSADP; the effect of mutations in R265[ECL3] impairing receptor activation is probably due to their effects on the conformational states of the receptor, rather than on agonist binding. Dinucleotide recognition at P2YRs is dependent on an optimal polyphosphate chain length and on precise substituents of both nucleoside moieties, suggesting that both pockets of the extended binding site participate in recognition. This is also consistent with dinucleotide docking to the AZD1283-bound $P2Y_{12}R$ structure, in which one nucleotide may bind in pocket 1 while the second half of the dinucleotide molecule is predicted to wrap around helix III to reach pocket 2, and the polyphosphate moiety occupies a highly cationic region of the binding site (Fig. 3c, d). The possibility of multiple binding pockets in the $P2Y_{12}R$ structure extends our knowledge of purinoceptor family ligand recognition mechanisms, and will serve as a template for designing improved orthosteric and allosteric drug candidates. The high conformational flexibility of the extracellular region suggests the potential for large conformational changes on nucleotide binding and activation, the understanding of which requires further structural studies of $P2Y_{12}R$ agonist complexes.

METHODS SUMMARY

The human $P2Y_{12}R$–BRIL fusion protein was expressed in Sf9 insect cells and purified by metal affinity chromatography. $P2Y_{12}R$–BRIL crystals were grown using a lipidic cubic phase (LCP) crystallization method. Diffraction data were collected from 15 crystals on the 23ID-B/D beamline (GM/CA CAT) at the Advanced Photon Source using a 10 μm minibeam at a wavelength of 1.0330 Å. The structure was solved by molecular replacement using the receptor portion of PAR1 (Protein Data Bank (PDB) accession 3VW7) and BRIL (PDB accession 1M6T) as initial models. The Schrödinger suite was used for docking, and saturation and competition assays were performed on cell membranes using [^{3}H]2MeSADP as a radiolabelled ligand.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 September 2013; accepted 27 January 2014.
Published online 23 March 2014.

1. Abbracchio, M. P. *et al.* International Union of Pharmacology LVIII: update on the P2Y G protein-coupled nucleotide receptors: from molecular mechanisms and pathophysiology to therapy. *Pharmacol. Rev.* **58,** 281–341 (2006).
2. Jacobson, K. A. & Boeynaems, J. M. P2Y nucleotide receptors: promise of therapeutic applications. *Drug Discov. Today* **15,** 570–578 (2010).
3. Bach, P. *et al.* Synthesis, structure–property relationships and pharmacokinetic evaluation of ethyl 6-aminonicotinate sulfonylureas as antagonists of the $P2Y_{12}$ receptor. *Eur. J. Med. Chem.* **65,** 360–375 (2013).
4. Jagroop, I. A., Burnstock, G. & Mikhailidis, D. P. Both the ADP receptors $P2Y_1$ and $P2Y_{12}$, play a role in controlling shape change in human platelets. *Platelets* **14,** 15–20 (2003).
5. Wallentin, L. *et al.* Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N. Engl. J. Med.* **361,** 1045–1057 (2009).
6. Husted, S. & van Giezen, J. J. Ticagrelor: the first reversibly binding oral $P2Y_{12}$ receptor antagonist. *Cardiovasc. Ther.* **27,** 259–274 (2009).
7. Savi, P. *et al.* The active metabolite of clopidogrel disrupts $P2Y_{12}$ receptor oligomers and partitions them out of lipid rafts. *Proc. Natl Acad. Sci. USA* **103,** 11069–11074 (2006).
8. Algaier, I., Jakubowski, J. A., Asai, F. & von Kügelgen, I. Interaction of the active metabolite of prasugrel, R-138727, with cysteine 97 and cysteine 175 of the human $P2Y_{12}$ receptor. *J. Thromb. Haemost.* **6,** 1908–1914 (2008).
9. Nylander, S., Mattsson, C., Ramstrom, S. & Lindahl, T. L. Synergistic action between inhibition of $P2Y_{12}$/$P2Y_1$ and $P2Y_{12}$/thrombin in ADP- and thrombin-induced human platelet activation. *Br. J. Pharmacol.* **142,** 1325–1331 (2004).
10. Bach, P. *et al.* Lead optimization of ethyl 6-aminonicotinate acyl sulfonamides as antagonists of the $P2Y_{12}$ receptor. Separation of the antithrombotic effect and bleeding for candidate drug AZD1283. *J. Med. Chem.* **56,** 7015–7024 (2013).
11. Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20,** 967–976 (2012).
12. Ballesteros, J. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25,** 366–428 (1995).
13. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63,** 1256–1272 (2003).
14. Hanson, M. A. *et al.* Crystal structure of a lipid G protein-coupled receptor. *Science* **335,** 851–855 (2012).
15. Zhang, C. *et al.* High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492,** 387–392 (2012).
16. Chee, M. J. *et al.* The third intracellular loop stabilizes the inactive state of the neuropeptide Y1 receptor. *J. Biol. Chem.* **283,** 33337–33346 (2008).
17. Schulz, A. & Schöneberg, T. The structural evolution of a P2Y-like G-protein-coupled receptor. *J. Biol. Chem.* **278,** 35531–35541 (2003).
18. Audet, M. & Bouvier, M. Restructuring G-protein- coupled receptor activation. *Cell* **151,** 14–23 (2012).
19. Rader, A. J. *et al.* Identification of core amino acids stabilizing rhodopsin. *Proc. Natl Acad. Sci. USA* **101,** 7246–7251 (2004).
20. Tan, Q. *et al.* Structure of the CCR5 chemokine receptor–HIV entry inhibitor maraviroc complex. *Science* **341,** 1387–1390 (2013).
21. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53,** 531–556 (2013).
22. Ding, Z., Bynagari, Y. S., Mada, S. R., Jakubowski, J. A. & Kunapuli, S. P. Studies on the role of the extracellular cysteines and oligomeric structures of the $P2Y_{12}$ receptor when interacting with antagonists. *J. Thromb. Haemost.* **7,** 232–234 (2009).
23. Ohlmann, P. *et al.* The platelet $P2Y_{12}$ receptor under normal and pathological conditions. Assessment with the radiolabeled selective antagonist [^{3}H]PSB-0413. *Purinergic Signal.* **9,** 59–66 (2013).
24. Cattaneo, M. The P2 receptors and congenital platelet function defects. *Semin. Thromb. Hemost.* **31,** 168–173 (2005).
25. Ignatovica, V., Megnis, K., Lapins, M., Schiöth, H. B. & Klovins, J. Identification and analysis of functionally important amino acids in human purinergic 12 receptor using a *Saccharomyces cerevisiae* expression system. *FEBS J.* **279,** 180–191 (2012).
26. Cattaneo, M. The platelet $P2Y_{12}$ receptor for adenosine diphosphate: congenital and drug-induced defects. *Blood* **117,** 2102–2112 (2011).
27. Cattaneo, M. *et al.* Molecular bases of defective signal transduction in the platelet $P2Y_{12}$ receptor of a patient with congenital bleeding. *Proc. Natl Acad. Sci. USA* **100,** 1978–1983 (2003).
28. Mao, Y., Zhang, L., Jin, J., Ashby, B. & Kunapuli, S. P. Mutational analysis of residues important for ligand interaction with the human $P2Y_{12}$ receptor. *Eur. J. Pharmacol.* **644,** 10–16 (2010).
29. Hasegawa, M. *et al.* Stereoselective inhibition of human platelet aggregation by R-138727, the active metabolite of CS-747 (prasugrel, LY640315), a novel $P2Y_{12}$ receptor inhibitor. *Thromb. Haemost.* **94,** 593–598 (2005).

Acknowledgements This work was supported by National Basic Research Program of China grants 2012CB910400 and 2012CB518000 (B.W., Q.Z.), National Institutes of Health (NIH) grants R01 AI100604 (B.W., Q.Z.) and U54 GM094618 (V.C., V.K., R.C.S.; Target GPCR-87), National Science Foundation of China grants 31370729 and 31170683 (B.W., Q.Z.), the National Institute of General Medical Sciences Postdoctoral Research Associate program (E.K.) and the NIH National Institute of Diabetes and Digestive and Kidney Diseases Intramural Research Program (K.A.J.). The authors thank AstraZeneca for their gift of AZD1283, and thank S. Nylander, F. Giordanetto and H. van Giezen for careful review and scientific feedback on the manuscript, A. Walker for assistance with manuscript preparation, and C. Wang and D. Wacker for help on collection of X-ray diffraction data.

Author Contributions K.Z. optimized the construct, expressed and purified human $P2Y_{12}R$–BRIL for crystallization, developed the purification procedure, performed crystallization trials and optimized crystallization conditions. J.Z. helped in construct and crystal optimization, and collected diffraction data. Z.-G.G. designed, performed and analysed ligand-binding and competition assays of wild-type and mutant $P2Y_{12}R$. D.Z. helped in expression and purification. L.Z. designed and made constructs for baculoviral expression. G.W.H. solved and refined the structure. S.M.M. performed and analysed ligand-binding assays. S.P. performed and analysed docking assays. E.K. helped in ligand synthesis of $P2Y_{12}R$. W.L. helped in crystal optimization. G.F. helped in crystallographic data collection. W.Z. developed the initial expression and purification protocol for $P2Y_{12}R$. C.E.M. provided compounds and discussed results. H.Y. helped to design and analysed docking assays. H.J. oversaw design and validation of $P2Y_{12}R$ models. V.C. helped to design and optimize LCP crystallization trials, collected and processed crystallographic data and wrote the manuscript. V.K. performed and analysed molecular modelling simulations, and wrote the manuscript. K.A.J. oversaw, designed and analysed ligand-binding assays and docking, and assisted with manuscript preparation. R.C.S. oversaw expression, purification and crystallization, and structure analysis/interpretation of $P2Y_{12}R$. B.W. and Q.Z. initiated the project, planned and analysed experiments, supervised the research and wrote the manuscript.

Author Information Atomic coordinates and structure factors have been deposited in the PDB under accession number 4NTJ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper.
Correspondence and requests for materials should be addressed to Q.Z. (zhaoq@simm.ac.cn).

METHODS

Protein engineering for structural studies. Human wild-type $P2Y_{12}R$ DNA (HUGO Gene Nomenclature Committee (HGNC) accession 18124) was codon optimized and synthesized by Genewiz for insect cell expression, and cloned into a modified pFastBac1 vector (Invitrogen) containing an expression cassette with a haemagglutinin (HA) signal sequence followed by a Flag tag at the N terminus and a PreScission protease site followed by a 10×His tag at the C terminus. To facilitate crystallogenesis, thermostabilized BRIL (PDB accession 1M6T) was fused into ICL3 of $P2Y_{12}R$ (T223–R224) with intact N and C termini excluding the start codon. The *$P2Y_{12}R$-BRIL* gene was further modified by introducing the D294$^{7.49}$N mutation based on the sequence alignment of conserved motifs.

Expression and purification of Sf9-expressed $P2Y_{12}R$ constructs for crystallization. High-titre recombinant baculovirus ($>10^8$ viral particles per ml) was obtained using the Bac-to-Bac Baculovirus Expression System (Invitrogen). Sf9 cells at a cell density of $2\text{–}3 \times 10^6$ cells ml^{-1} were infected with virus at a multiplicity of infection (m.o.i.) of 5. Cells were harvested by centrifugation at 48 h post-infection and stored at −80 °C until use. Insect cell membranes were disrupted by thawing frozen cell pellets in a hypotonic buffer containing 10 mM HEPES, pH 7.5, 10 mM $MgCl_2$, 20 mM KCl and protease inhibitor cocktail (Roche) with the ratio of 1 tablet per 100 ml lysis buffer. Extensive washing of the raw membranes was performed by repeated centrifugation in the same buffer and then in a high salt buffer containing 50 mM HEPES, pH 7.5, 10 mM $MgCl_2$, 20 mM KCl and 1 M NaCl (three times each).

Purified membranes were thawed on ice in the presence of 200 μM AZD1283, 2 mg ml^{-1} iodoacetamide, and EDTA-free protease inhibitor cocktail (Roche), and incubated at 4 °C for 30 min before solubilization. $P2Y_{12}R$-BRIL was extracted from the membrane by adding *n*-dodecyl-β-D-maltopyranoside (DDM; Affymetrix) and cholesteryl hemisuccinate (CHS; Sigma) to the membrane solution to a final concentration of 0.5% (w/v) and 0.1% (w/v), respectively, and stirring was continued at 4 °C for 2.5 h. The supernatant was isolated by centrifugation at 160,000*g* for 30 min and incubated with TALON IMAC resin (Clontech) overnight at 4 °C. The resin was then washed with twenty column volumes of 50 mM HEPES, pH 7.5, 1 M NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS and 20 mM imidazole. The protein was then eluted with 5 column volumes of 50 mM HEPES, pH 7.5, 1 M NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS, 300 mM imidazole and 200 μM AZD1283. A PD MiniTrap G-25 column (GE Healthcare) was used to remove imidazole. The protein was then treated overnight with His-tagged PreScission protease (20 μg per 500 ml of expressed material) and His-tagged PNGase F (20 μg per 500 ml of expressed material) to remove the C-terminal His tag and deglycosylate the receptor. PreScission protease, PNGase F and the cleaved 10×His tag were removed from the sample by passing the sample over Ni-NTA superflow resin (Qiagen). The receptor was then concentrated to 20–30 mg ml^{-1} with a 100 kDa molecular weight cut-off concentrator (Millipore). Protein purity and monodispersity was tested by SDS–PAGE and aSEC. Typically, the protein purity exceeded 95% and the aSEC profile showed a single peak, indicative of receptor monodispersity.

For aSEC analysis of $P2Y_{12}R$ in complex with R-138727 (Alsachim), the receptor was first treated with 100 μM R-138727 on insect cell membrane at 4 °C for 1 h and then purified under a similar protocol without further supplement with ligand thereafter.

Lipidic cubic phase crystallization of $P2Y_{12}R$. The $P2Y_{12}R$–BRIL construct was crystallized using the lipidic cubic phase (LCP) method by mixing 40% of ~40 mg ml^{-1} protein with 60% lipid (monoolein and cholesterol 10:1 by mass) using a syringe lipid mixer as described previously[30]. After a clear LCP formed, the mixture was dispensed onto glass sandwich plates (Shanghai FAstal BioTech) into 40 nl drops and overlaid with 800 nl precipitant solution using a Mosquito LCP robot (TTP LabTech). Crystals appeared after 3 days and reached their full size within 2 weeks in 0.05–0.15 M ammonium formate, 0.1 M sodium cacodylate, pH 6.0–6.5, 25–35% PEG400 and 200 μM AZD1283. Crystals were harvested directly from LCP using 100–150 μm micro-loops (M2-L19-100/150, MiTeGen) and flash frozen in liquid nitrogen.

Data collection and structure solution. X-ray data were collected on the 23ID-B/D beamline (GM/CA CAT) at the Advanced Photon Source using a 10 μm mini-beam (at a wavelength of 1.0330 Å) and a MarMosaic 300 CCD detector. Among the crystal samples screened, most crystals diffracted to 3.0–2.6 Å resolution when exposed to 1 s of unattenuated beam using 1° oscillation. Data from the 15 best-diffracting crystals were integrated and scaled to an overall 2.6 Å resolution using HKL2000[31]. Initial phase information was obtained by molecular replacement using the receptor portion of PAR1 (PDB accession 3VW7) and BRIL (PDB accession 1M6T) independently with the program Phaser[32]. All refinements were performed with Refmac5[33] and Buster[34] followed by manual examination and rebuilding of the refined coordinates in the program Coot[35] using both $2mF_o - DF_c$ and $mF_o - DF_c$ maps.

Ligand-binding assays. Wild-type and mutant $P2Y_{12}R$ plasmids with single amino acid substitutions (Extended Data Table 3) were cloned into PCDNA3.0 and transfected into COS7 cells using Lipofectamine 2000 (Life Technologies). *Sf*9 cells were harvested 48 h after transfection. After harvesting, cells were homogenized for 15 s and then centrifuged for 10 min at 1,000*g*. The suspension was re-centrifuged at 20,000*g* for 60 min. The resulting pellet was re-suspended, homogenized, split into aliquots and maintained at −80 °C in a freezer until use. Protein concentrations were measured using Bio-Rad protein assay reagents. Membranes for binding with the constructs containing BRIL and the point mutation (Extended Data Table 2) were prepared following the same procedure using Sf9 cells.

For saturation experiments, 50 μl [^{3}H]2MeSADP (3.5 Ci $mmol^{-1}$, from 0.4 to 46 nM; Moravek) was incubated with 100 μl wild-type and mutant $P2Y_{12}R$ membrane preparations (5 μg per tube) in a total assay volume of 200 μl Tris-HCl buffer containing 10 mM $MgCl_2$. AZD1283 (10 μM) was used to determine non-specific binding. For displacement experiments, increasing concentrations of AZD1283 were incubated with wild-type or mutant membrane preparations (5–10 μg) and [^{3}H]2MeSADP (10 nM) at 25 °C for 30 min. ADP, 2MeSADP and 2MeSATP were obtained from Sigma PSB-0739 was obtained from Tocris. The reaction was terminated by harvesting with a 24-channnel Brandel cell harvester and followed by washing twice with 5 ml cold Tris-HCl buffer containing 10 mM $MgCl_2$. Radioactivity was measured using a scintillation counter (Tri-Carb 2810TR). Data were analysed using Prism 6 (GraphPad).

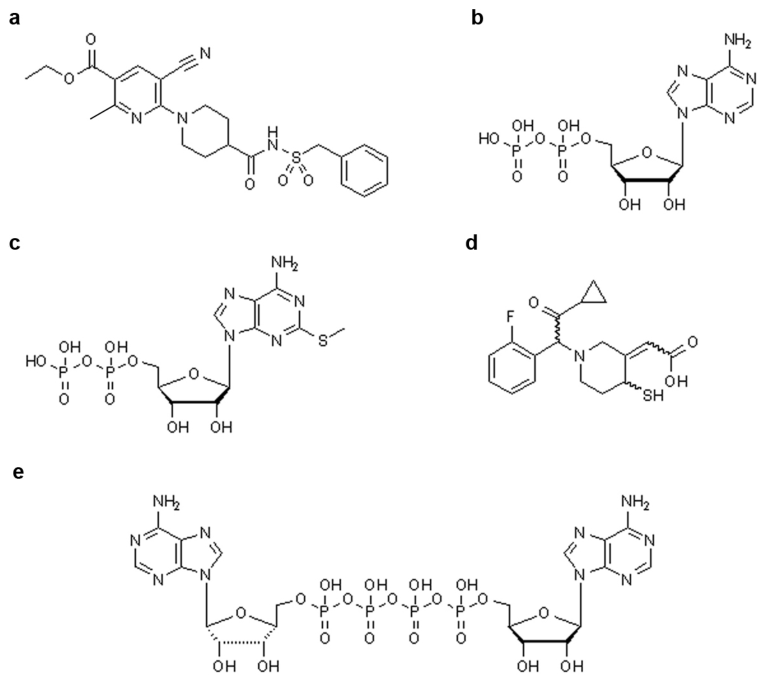
Docking and molecular modelling. The $P2Y_{12}R$ structure was prepared using the Protein Preparation Wizard[36] tool implemented in the Schrödinger suite, adding all the hydrogen atoms and the missing side chains of residues whose backbone coordinates were observed in the structure. The orientation of polar hydrogens was optimized, the protein protonation states were adjusted and the overall structure was minimized with harmonic restraints on the heavy atoms, to remove strain. Then, all the hetero groups and water molecules were deleted.

The SiteMap tool of the Schrödinger suite was used to identify potential binding sites in the structure. A bifurcated cavity was identified on the extracellular side of the receptor and was selected as the docking site. Molecular docking of selected compounds (ADP, 2MeSADP, AZD1283, Ap4A and the active metabolites of clopidogrel and prasugrel) at the $P2Y_{12}R$ structure was performed by means of the Glide package from the Schrödinger suite. In particular, a Glide Grid was centred on the centroid of residues located within 6 Å from the previously identified cavity (considering both pocket 1 and pocket 2). The Glide Grid was built using an inner box (ligand diameter midpoint box) of 14 Å × 14 Å × 14 Å (so that both pockets could be explored) and an outer box that extended 10 Å in each direction from the inner one (so that ligands up to 20 Å could be docked). Docking of ligands was performed in the rigid binding site using the standard precision procedure. The top scoring docking conformations for each ligand were subjected to visual inspection and analysis of protein–ligand interactions to select the final binding conformations in agreement with the experimental data.

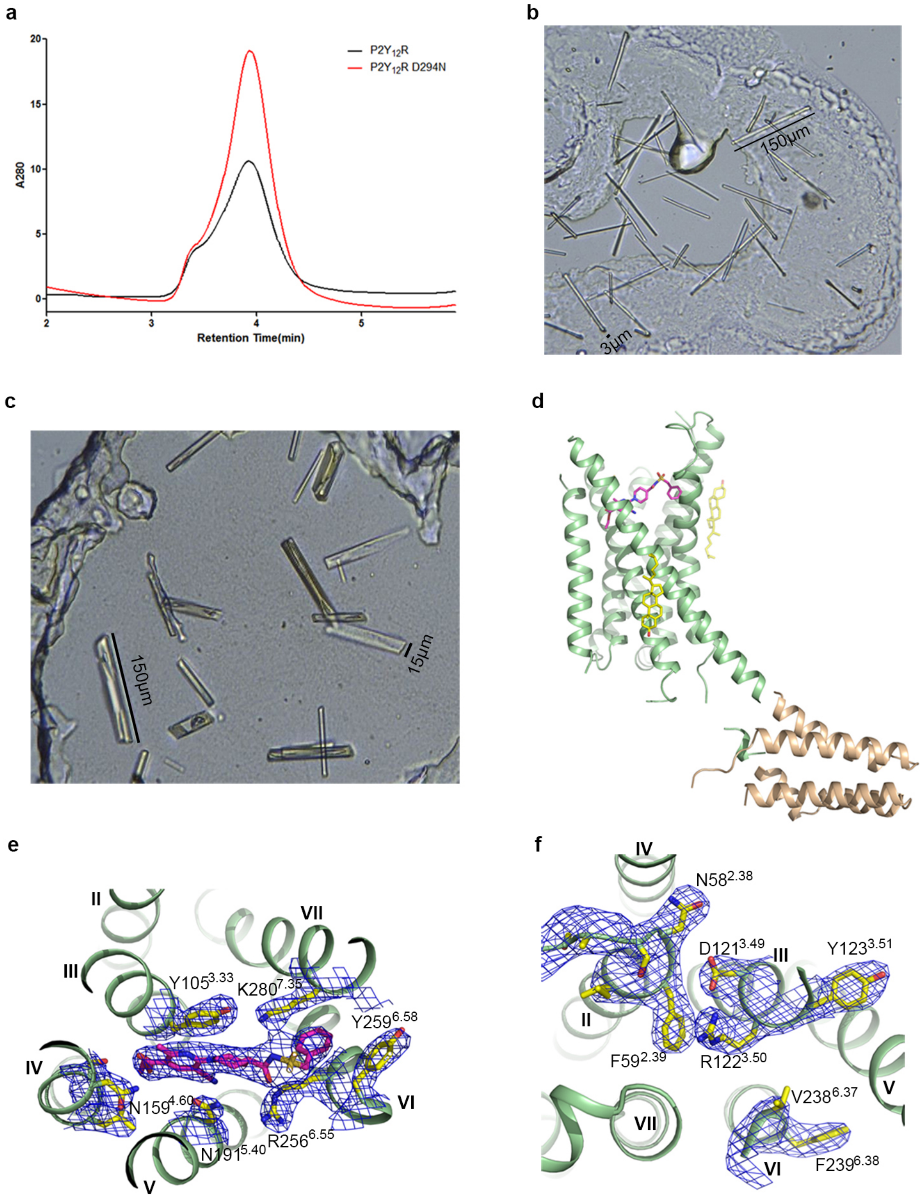
DNA sequence of the crystallization construct. ATGAAGACGATCATCGCCC
TGAGCTACATCTTCTGCCTGGTGTTCGCCGACTACAAGGACGATGATG
ACGGCGCGCCGCAAGCCGTGGACAACCTCACATCAGCCCCTGGCAACA
CCTCCCTCTGTACCCGCGACTACAAGATCACACAAGTTCTCTTCCCCCT
CCTCTACACAGTGTTGTTCTTCGTCGGCCTCATCACCAACGGATTGGCT
ATGCGTATCTTCTTCCAGATCCGCTCCAAGTCTAACTTCATCATCTTCC
TGAAGAACACTGTGATCTCGGACCTGCTCATGATCCTCACATTCCCATT
CAAGATCCTGTCAGATGCCAAGCTCGGTACTGGCCCGTTGCGTACATT
CGTCTGCCAGGTTACCTCTGTGATCTTCTACTTCACTATGTACATCAGC
ATCTCATTCCTGGGTCTCATCACCATCGACAGGTACCAAAAGACCACTA
GACCCTTCAAGACTAGCAACCCTAAGAACTTGCTGGGCGCTAAGATCC
TGAGCGTGGTCATCTGGGCCTTCATGTTCCTCTTGTCACTGCCCAACAT
GATCCTCACCAACAGGCAGCCTAGAGATAAGAACGTGAAGAAGTGTTC
ATTCCTCAAGTCGGAGTTCGGATTGGTTTGGCACGAAATCGTGAACTA
CATCTGCCAAGTCATCTTCTGGATCAACTTCCTGATCGTTATCGTGTGT
TACACATTGATCACCAAGGAGCTCTACAGGTCCTACGTCCGTACTGCT
GATCTGGAAGACAATTGGGAAACTCTGAACGACAATCTCAAGGTGATC
GAGAAGGCTGACAATGCTGCACAAGTCAAAGACGCTCTGACCAAGAT
GAGGGCAGCAGCCCTGGACGCTCAGAAGGCCACTCCACCTAAGCTCGA
GGACAAGAGCCCAGATAGCCCTGAAATGAAAGACTTTCGGCATGGAT
TCGACATTCTGGTGGGACAGATTGATGATGCACTCAAGCTGGCCAATG
AAGGGAAAGTCAAGGAAGCACAAGCAGCCGCTGAGCAGCTGAAGACC
ACCCGGAATGCATACATTCAGAAGTACCTGCGCGGAGTCGGCAAGGTT
CCTAGGAAGAAGGTCAACGTTAAGGTGTTCATCATCATCGCTGTCTTC
TTCATCTGCTTCGTTCCATTCCACTTCGCCCGTATCCCGTACACTTTGT
CCCAAACACGCGACGTGTTCGATTGTACCGCTGAGAACACTCTGTTCT
ACGTCAAGGAATCCACATTGTGGCTGACCTCTCTGAACGCCTGCCTCA
ACCCATTCATCTACTTCTTCCTCTGTAAGTCTTTCCGCAACTCGTTGAT
CTCCATGCTGAAGTGCCCTAACTCTGCTACCAGCCTGTCCCAAGATAA

CAGAAAGAAGGAGCAGGACGGAGGCGACCCGAACGAGGAAACCCCGATGGGCCGGCCTCTGGAAGTTCTGTTCCAGGGGCCCCATCATCATCATCATCATCATCATCATCATTAG.

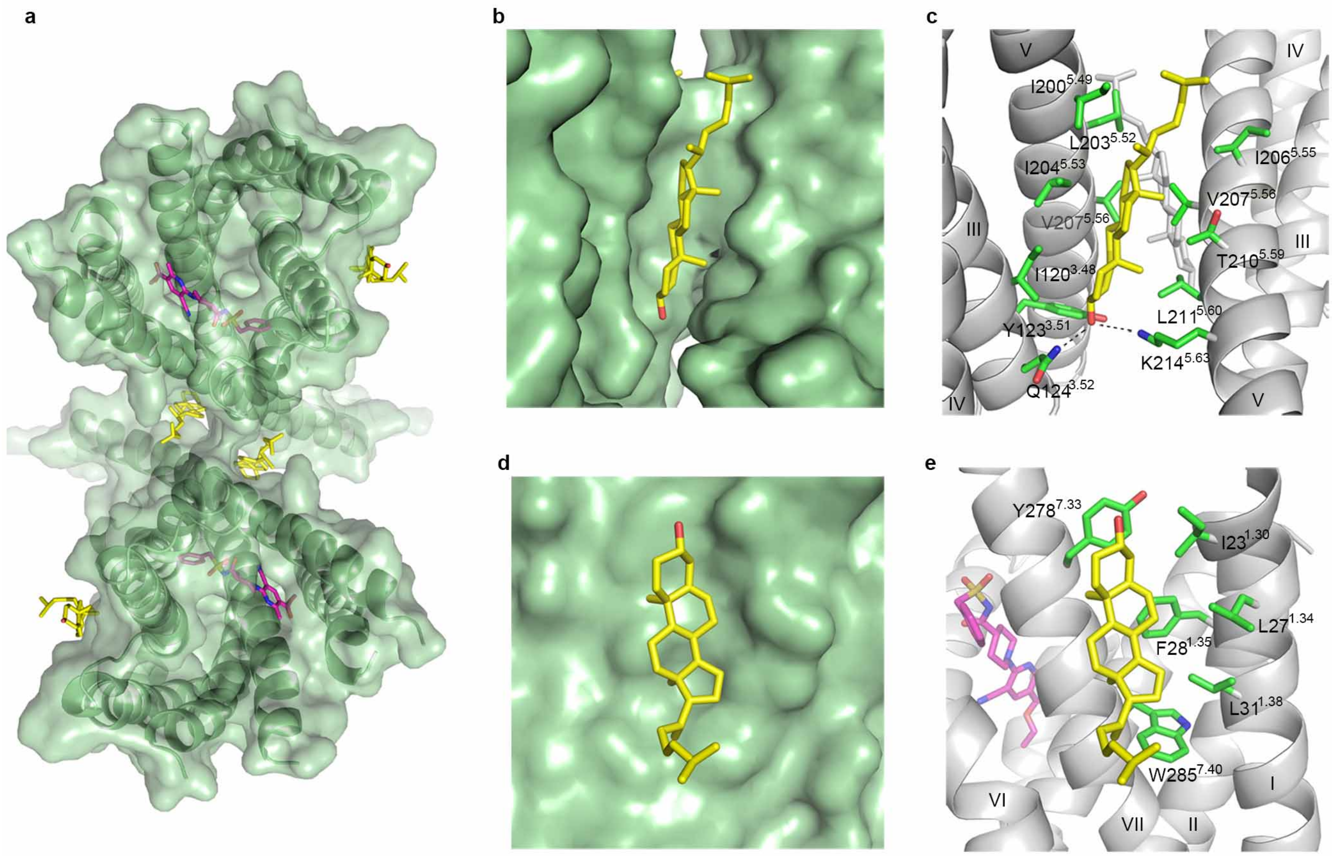
30. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4,** 706–731 (2009).
31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
33. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60,** 2184–2195 (2004).
34. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target–structure restraints in *BUSTER*. *Acta Crystallogr. D* **68,** 368–380 (2012).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
36. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27,** 221–234 (2013).



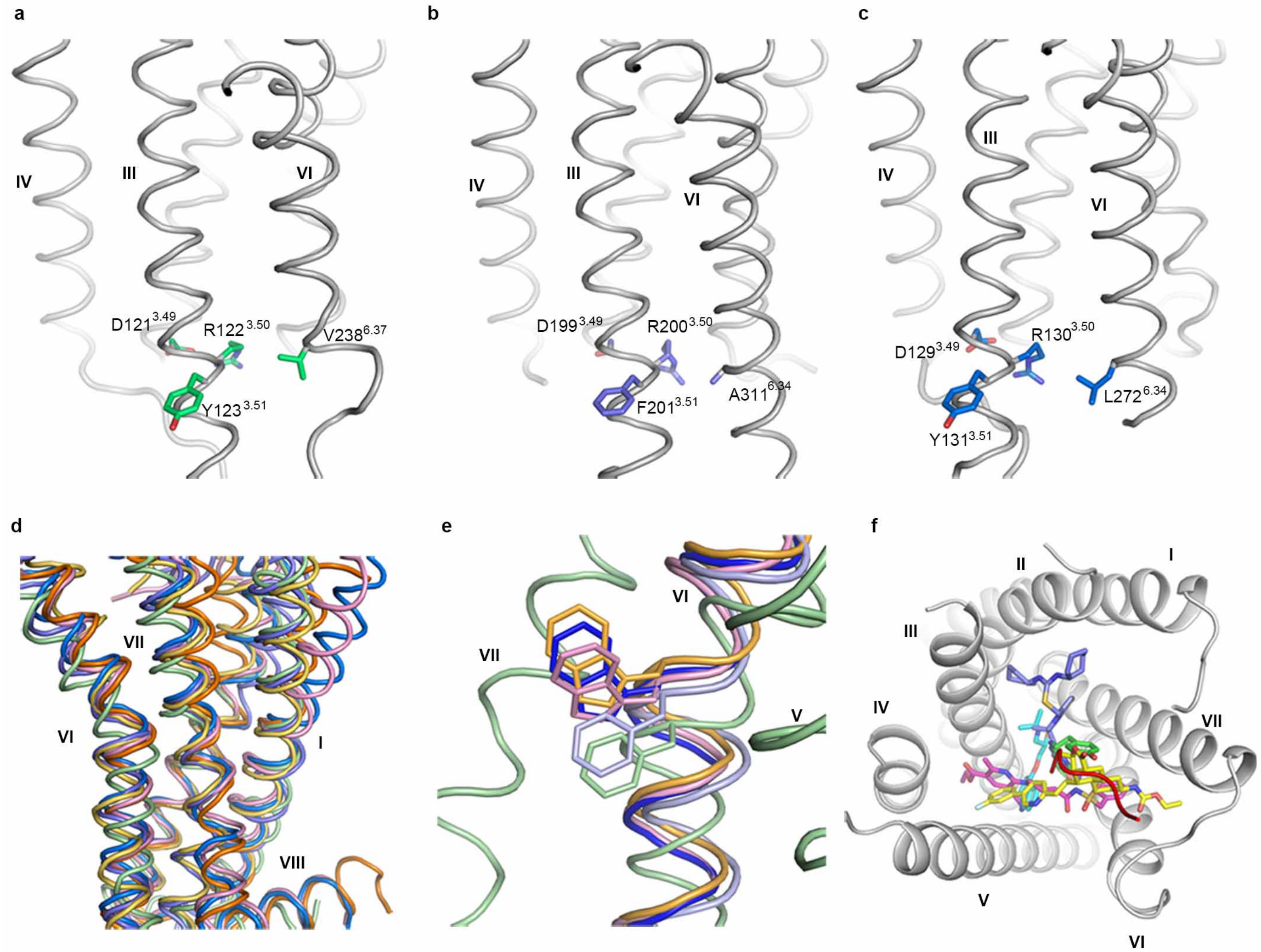
Extended Data Figure 1 | Chemical structures of different P2Y$_{12}$R ligands. **a**, AZD1283. **b**, ADP. **c**, 2MeSADP. **d**, R-138727. **e**, Diadenosine tetraphosphate (Ap4A).



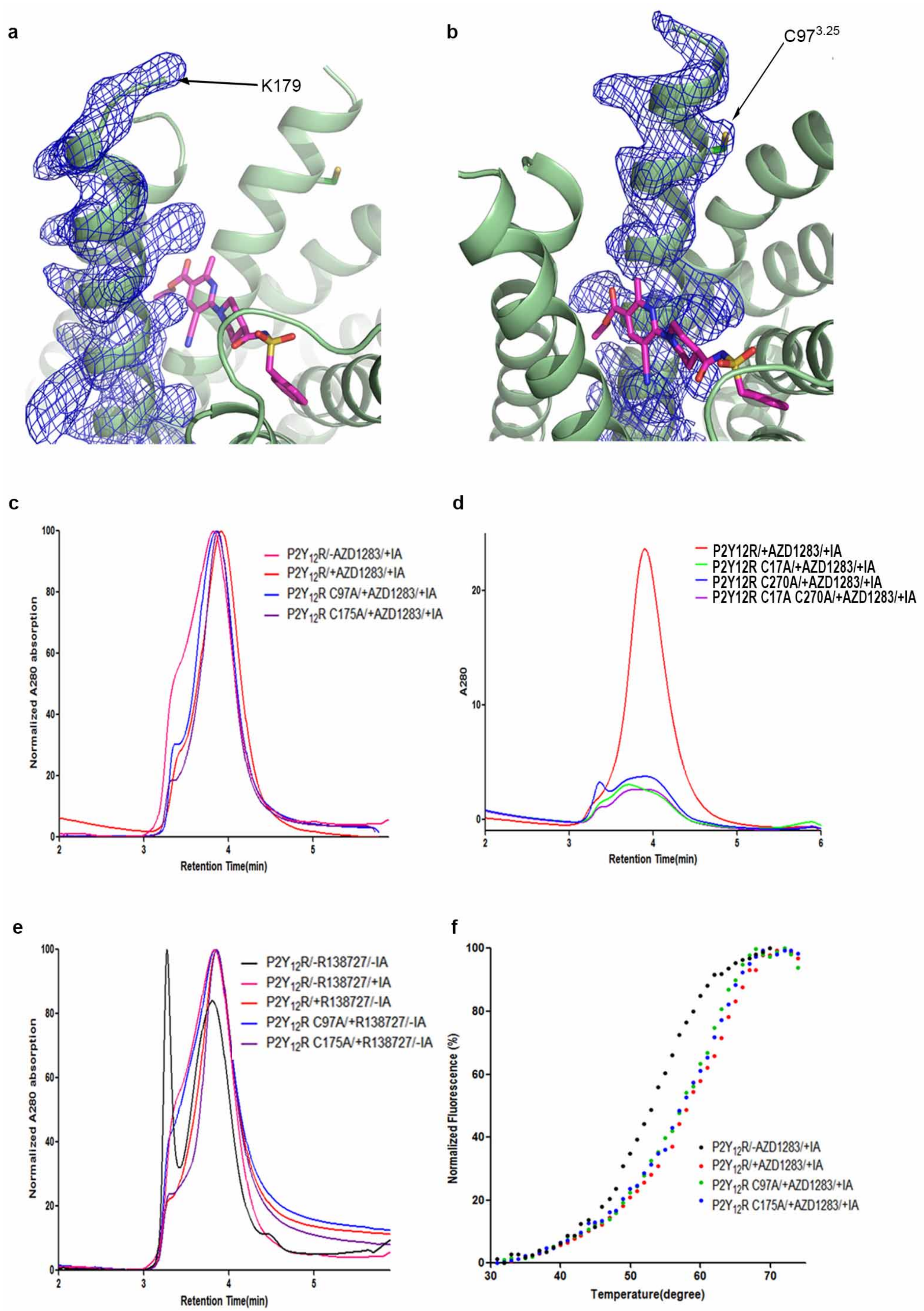
Extended Data Figure 2 | Size-exclusion chromatography traces, crystals and overall structure of P2Y$_{12}$R–AZD1283 complex. **a**, aSEC traces of P2Y$_{12}$R–BRIL (black) and P2Y$_{12}$R(D294N)–BRIL (red) purified in complex with AZD1283. The samples are expressed and purified in parallel from roughly the same amount of cells. **b**, Crystals of P2Y$_{12}$R–BRIL and AZD1283 complex. The size of crystals is roughly $150 \times 5 \times 5$ µm. **c**, Crystals of P2Y$_{12}$R(D294N)–BRIL and AZD1283 complex. The size of crystals is roughly $150 \times 15 \times 15$ µm. **d**, Cartoon representation of P2Y$_{12}$R(D294N)–BRIL. The P2Y$_{12}$R is shown in pale green ribbons, BRIL is in wheat ribbons, AZD1283 is magenta carbons, and cholesterols and lipids are yellow carbons. **e**, The $2mF_o - DF_c$ electron density map of the ligand-binding pocket contoured at 1.2σ. **f**, The $2mF_o - DF_c$ electron density map of the DRY motif region contoured at 1.2σ.



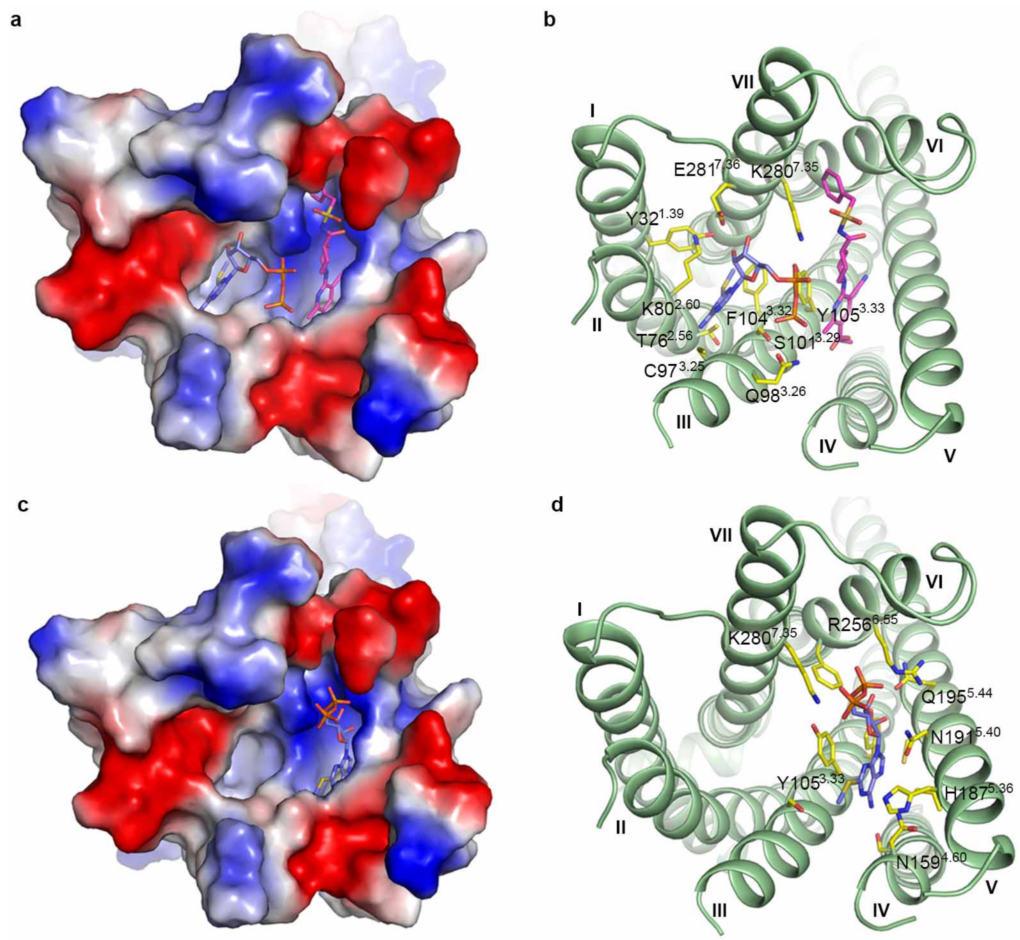
Extended Data Figure 3 | Dimeric receptor association and interactions with cholesterols in the crystals of P2Y$_{12}$R. **a**, Two P2Y$_{12}$R molecules make contact with each other, as mediated by helix V and two molecules of cholesterol related by a two-fold axis. **a**, **b**, The detailed interactions of the cholesterol molecules on helices III and V are shown in surface (**b**) and in cartoon (**c**) representations. The cholesterol molecule is coloured in yellow carbons, and P2Y$_{12}$R is shown in pale green (surface) or grey (cartoon). Residues within 4 Å of cholesterols are represented as green sticks. Hydrogen-bond interactions are indicated by dashed lines. As the interactions of the two cholesterols at the interface are identical, only one is revealed in detail. **d**, **e**, The binding site of cholesterol between helices I and VII is shown in surface (**d**) and in cartoon (**e**) representation. AZD1283 is shown in magenta carbons.



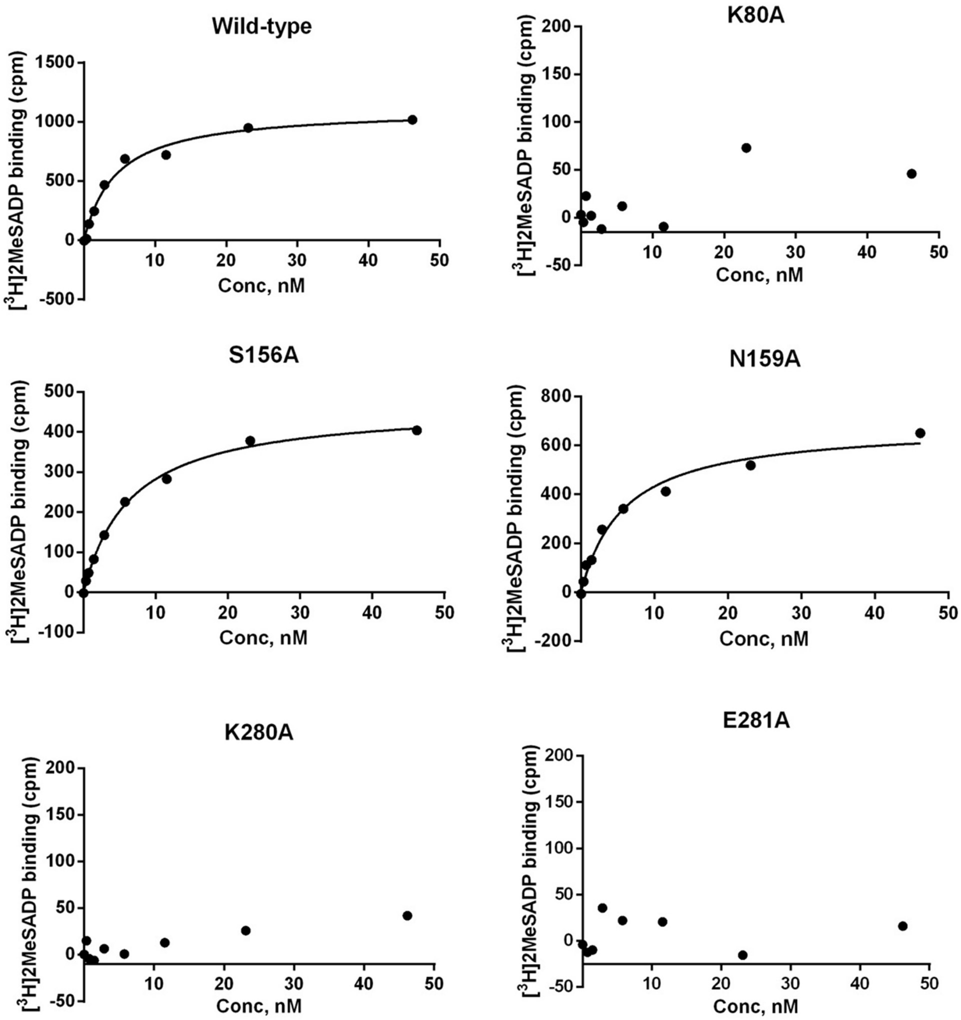
Extended Data Figure 4 | Comparison of relative residue positions between helix III and helix VI of P2Y$_{12}$R, PAR1 and β$_2$AR. **a**, P2Y$_{12}$R. **b**, PAR1. **c**, β$_2$AR. The receptors are shown in grey ribbon representation. The DR$^{3.50}$Y(F) motif and corresponding 6.37 (or 6.34) positions are shown as sticks of green, slate and blue carbons, respectively. **d**, Superimposition of P2Y$_{12}$R with other GPCR structures. P2Y$_{12}$R (pale green), β$_2$AR (PBD accession 2RH1; blue), A$_{2A}$AR (PDB accession 3EML; orange), κ-OR (PDB accession 4DJH; pink), NTSR1 (PDB accession 4GRV; yellow) and PAR1 (PDB accession 3VW7; slate) are superimposed and shown as ribbons. Transmembrane helices I, II and VII overlay relatively well, whereas the position of helix VI is substantially different in P2Y$_{12}$R. **e**, Comparison of W(F)$^{6.48}$ positions in P2Y$_{12}$R (pale green), β$_2$AR (blue), A$_{2A}$AR (orange), κ-OR (pink) and PAR1 (light blue). For receptors other than P2Y$_{12}$R, only helix VI is shown, and the residues at position 6.48 are shown as sticks. **f**, The comparison of ligand-binding sites of GPCRs from α (β$_2$AR and A$_{2A}$AR), β (NTSR1), γ (CXCR4; PDB accession 3OE0) and δ (P2Y$_{12}$R and PAR1) subgroups. The structure of P2Y$_{12}$R is shown in grey cartoon representation and AZD1283 is shown as magenta sticks. The ligands from other receptors (carazolol, cyan; ZM241385, green; neurotensin, red cartoon; IT1t, purple; vorapaxar, yellow) are placed at their corresponding positions in the 7TM bundle.


a
K179
b
C97[3.25]
c
P2Y12R/-AZD1283/+IA
P2Y12R/+AZD1283/+IA
P2Y12R C97A/+AZD1283/+IA
P2Y12R C175A/+AZD1283/+IA
Normalized A280 absorption
Retention Time(min)
d
P2Y12R/+AZD1283/+IA
P2Y12R C17A/+AZD1283/+IA
P2Y12R C270A/+AZD1283/+IA
P2Y12R C17A C270A/+AZD1283/+IA
A280
Retention Time(min)
e
P2Y12R/-R138727/-IA
P2Y12R/-R138727/+IA
P2Y12R/+R138727/-IA
P2Y12R C97A/+R138727/-IA
P2Y12R C175A/+R138727/-IA
Normalized A280 absorption
Retention Time(min)
f
P2Y12R/-AZD1283/+IA
P2Y12R/+AZD1283/+IA
P2Y12R C97A/+AZD1283/+IA
P2Y12R C175A/+AZD1283/+IA
Normalized Fluorescence (%)
Temperature(degree)

Extended Data Figure 5 | The conserved helix III–ECL2 disulphide bond might be dynamic in AZD1283-bound $P2Y_{12}R$. **a**, **b**, The electron density of helix V (**a**) and helix III (**b**). Electron density is represented by a $2mF_o - DF_c$ map countered at 1.2σ. The K179 and $C97^{3.25}$ side chains are indicated with black arrows. **c**–**f**, Effects of cysteine mutations on $P2Y_{12}R$–BRIL. As the stability of wild-type $P2Y_{12}R$ is very poor, all the constructs here contain fusion protein and D294N mutation as described in the crystallization of the receptor. **c**, Comparison of aSEC traces of $P2Y_{12}R$ apo and $P2Y_{12}R$, C97A, C175A in complex with AZD1283. The samples were treated with iodoacetamide (IA) before extraction. **d**, Comparison of aSEC traces of $P2Y_{12}R$, C17A, C270A and C17A/C270A in complex with AZD1283. **e**, aSEC curves of purified $P2Y_{12}R$ (red), $P2Y_{12}R$ C97A (blue) and $P2Y_{12}R$ C175A (purple) with R-138727, the active metabolite of prasugrel, which binds irreversibly to $P2Y_{12}R$ by interacting with its cysteine residue(s). aSEC curve of apo $P2Y_{12}R$ is shown in black. It is obvious that treatment of R-138727 greatly improves the homogeneity of $P2Y_{12}R$, which is not affected by the C175A mutation. However, the C97A mutation almost completely abolishes the effect of R-138727, indicating that $C97^{3.25}$ is the binding site of this compound. **f**, The melting curves of $P2Y_{12}R$, C97A, C175A binding with AZD1283 and $P2Y_{12}R$ in the apo form.



Extended Data Figure 6 | The hypothetical binding modes of 2MeSADP to antagonist-bound state P2Y$_{12}$R. **a**, **b**, An electrostatics surface representation (**a**) and a cartoon representation (**b**) of the hypothetical docking model of P2Y$_{12}$R with bound 2MeSADP in pocket 2. The agonist 2MeSADP is shown as sticks (deep blue carbons). The side chains of residues that are involved in the binding of 2MeSADP are also labelled and shown as sticks (yellow carbons). **c**, **d**, An electrostatics surface representation (**c**) and a cartoon representation (**d**) of the hypothetical docking model of P2Y$_{12}$R with bound 2MeSADP in pocket 1.



Extended Data Figure 7 | Representative saturation curves of [^{3}H]2MeSADP-specific binding to wild-type P2Y$_{12}$R and various mutant receptors at a concentration range of 0.4–46 nM. The calculated K_d values from 3–6 independent experiments are listed in Extended Data Table 3. Non-specific binding was determined using 10 μM AZD1283.

Extended Data Table 1 | Data collection and refinement statistics

Data Collection*	
Space group	*C2*
Cell dimensions	
a, b, c (Å)	98.6, 156.4, 47.8
α, β, γ (°)	90.0, 111.1, 90.0
Number of reflections processed	73,502
Number of unique reflections	19,116
Resolution (Å)	50.0-2.6 (2.7-2.6) †
R_{merge} (%)	10.3 (95.8)
$CC_{1/2}$	0.99 (0.64)
Mean $I/\sigma(I)$	14.4 (1.2)
Completeness (%)	94.1 (79.5)
Redundancy	3.8 (3.1)
Refinement	
Resolution (Å)	50.0-2.6
Number of reflections (test set)	19,094 (990)
R_{work} / R_{free} (%)	22.0 / 24.6
Number of atoms	
Protein	2,886
Ligand	33
Cholesterol	56
Lipids, PEG and waters	50
Overall *B* values (Å^2)	
$P2Y_{12}R$	107.4
BRIL	99.7
Ligand	95.6
Cholesterol	133.2
Lipids and waters	119.1
RMSD	
Bond lengths (Å)	0.010
Bond angles (°)	1.05
Ramachandran plot statistics (%) ‡	
Favored regions	96.4
Allowed regions	3.6
Disallowed regions	0.0

* Fifteen crystals were used for data processing.
† Values in parentheses are for highest-resolution shell.
‡ As defined in MolProbity.

Extended Data Table 2 | Binding affinities for different $P2Y_{12}R$ constructs

Inhibition of binding	K_i (nM), WT	K_i (nM), WT+BRIL	K_i (nM), WT+D294N mutation	K_i (nM), WT+BRIL +D294N mutation
ADP	332±35	405±122	169±81	280±85
2MeSADP	120±7.3	91.3±30.7	80.8±16.8	128±48
2MeSATP	119±19.8	68.0±20.8	82.4±28.5	93.2±19.9
AZD1283	31.5±11.9	39.3±9.9	42.2±27.7	43.9±10.0
PSB-0739	275±58	155±73	72.8±16.7 †	157±58
Ticagrelor	454±110	299±138	424±193	182±70
Saturation binding	**WT**	**WT+BRIL**	**WT+D294N mutation**	**WT+BRIL +D294N mutation**
K_d (nM)	35.9±6.3	36.7±3.6	33.6±10.2	51.5±12.0
B_{max} (pmol/mg protein)	6.34±0.27	6.87±0.78	11.4±1.1	11.9±0.63

Data represent mean ± standard error of the mean (s.e.m.) from 3–8 separate experiments performed in duplicate.
† One-way analysis of variance with post-hoc test yielded significant difference ($P < 0.05$) compared with wild type (WT) but not with WT+BRIL and WT+BRIL+D294N groups ($P > 0.05$). The affinities of other ligands at wild type are not significantly different from their affinities at mutants ($P > 0.05$).

Extended Data Table 3 | [^{3}H]2MeSADP binding affinities for different P2Y$_{12}$R mutants

Constructs	[^{3}H]2MeSADP (K_d, nM)
WT	4.25 ± 0.87
K80A	N.S.*
S156A	5.66 ± 0.72
N159A	6.39 ± 0.94
K280A	N.S.
E281A	N.S.

* N.S., not saturable or negligible specific binding within the radioligand concentrations used (0.4–46 nM). Results are expressed as mean ± s.e.m. from 3–6 independent experiments performed in duplicate.

doi:10.1038/nature13288

Agonist-bound structure of the human $P2Y_{12}$ receptor

Jin Zhang[1]*, Kaihua Zhang[1]*, Zhan-Guo Gao[2], Silvia Paoletta[2], Dandan Zhang[1], Gye Won Han[3], Tingting Li[1], Limin Ma[1], Wenru Zhang[1], Christa E. Müller[4], Huaiyu Yang[5], Hualiang Jiang[5], Vadim Cherezov[3], Vsevolod Katritch[3], Kenneth A. Jacobson[2], Raymond C. Stevens[3,6], Beili Wu[1] & Qiang Zhao[1]

The $P2Y_{12}$ receptor ($P2Y_{12}R$), one of eight members of the P2YR family expressed in humans, is one of the most prominent clinical drug targets for inhibition of platelet aggregation. Although mutagenesis and modelling studies of the $P2Y_{12}R$ provided useful insights into ligand binding[1–4], the agonist and antagonist recognition and function at the $P2Y_{12}R$ remain poorly understood at the molecular level. Here we report the structures of the human $P2Y_{12}R$ in complex with the full agonist 2-methylthio-adenosine-5′-diphosphate (2MeSADP, a close analogue of endogenous agonist ADP) at 2.5 Å resolution, and the corresponding ATP derivative 2-methylthio-adenosine-5′-triphosphate (2MeSATP) at 3.1 Å resolution. These structures, together with the structure of the $P2Y_{12}R$ with antagonist ethyl 6-(4-((benzylsulfonyl) carbamoyl)piperidin-1-yl)-5-cyano-2-methylnicotinate (AZD1283)[5], reveal striking conformational changes between nucleotide and non-nucleotide ligand complexes in the extracellular regions. Further analysis of these changes provides insight into a distinct ligand binding landscape in the δ-group of class A G-protein-coupled receptors (GPCRs). Agonist and non-nucleotide antagonist adopt different orientations in the $P2Y_{12}R$, with only partially overlapped binding pockets. The agonist-bound $P2Y_{12}R$ structure answers long-standing questions surrounding $P2Y_{12}R$–agonist recognition, and reveals interactions with several residues that had not been reported to be involved in agonist binding. As a first example, to our knowledge, of a GPCR in which agonist access to the binding pocket requires large-scale rearrangements in the highly malleable extracellular region, the structural and docking studies will therefore provide invaluable insight into the pharmacology and mechanisms of action of agonists and different classes of antagonists for the $P2Y_{12}R$ and potentially for other closely related P2YRs.

After sensing their endogenous extracellular ligands, GPCRs activate associated intracellular signal transduction pathways that subsequently lead to physiological responses[6]. Structures of five GPCRs (rhodopsin[7,8], β_1 (β_1AR)[9,10] and β_2 adrenergic receptors (β_2AR)[11,12], A_{2A} adenosine receptor (A_{2A}AR)[13,14] and M2 muscarinic receptor[15,16]) have now been determined in both antagonist- and agonist-bound states. Each of these five receptors, however, belong to the α-group of class A GPCRs[17]. Here we describe a 2.5 Å structure of human $P2Y_{12}R$ bound to the full agonist 2MeSADP, and a 3.1 Å structure of $P2Y_{12}R$ bound to a potential partial agonist 2MeSATP (Extended Data Table 1). Together with the 2.7 Å structure of $P2Y_{12}R$ bound to the antagonist AZD1283 reported in the accompanying paper[5], this allows the first crystallographic assessment of a receptor with both agonist- and antagonist-bound structures in the δ-group of class A GPCRs.

All three $P2Y_{12}R$ structures were determined using the same thermostabilizing construct[5]. The receptor conformations of the 2MeSADP and 2MeSATP bound complexes are very similar (Cα root mean squared deviation (r.m.s.d.) = 0.6 Å), and both ligands overlay well in the same binding pocket (r.m.s.d. of common atoms = 0.6 Å), with the γ-phosphate group of 2MeSATP extended towards the extracellular surface (Fig. 1 and Extended Data Fig. 1). As the two structures are similar, we will focus our discussions mainly on the higher resolution $P2Y_{12}R$–2MeSADP structure, mentioning the specific differences of the 2MeSATP-bound complex when relevant.

The receptor is folded into a canonical seven transmembrane (7TM) bundle with a partially resolved intracellular helix VIII (Fig. 1). The straight conformation and tilted orientation of helix V observed in the AZD1283-bound structure[5] is probably a genuine structural feature inherent to the $P2Y_{12}R$, as it is consistent in all three structures despite the different crystal packing configurations (Extended Data Fig. 2). Both disulphide bonds, bridging the amino terminus (C17) with helix VII ($C270^{7.25}$, superscript indicates Ballesteros–Weinstein residue numbering[18]) and the highly conserved disulphide bond between helix III ($C97^{3.25}$) and extracellular loop 2 (ECL2, C175) are observed.

Comparison of the $P2Y_{12}R$–AZD1283 and $P2Y_{12}R$–2MeSADP complex structures reveals remarkable differences. The largest conformational change occurs in helices VI and VII. The extracellular part of helix VI in the 2MeSADP-bound structure shifts over 10 Å and helix VII over 5 Å towards the axis of the 7TM helical bundle, as compared to the antagonist-bound structure (Fig. 2). In the AZD1283-bound structure, the phenyl moiety of the antagonist apparently prevents this inward movement of helix VI, which may explain the difference between the two complexes. The inward shift of helices VI and VII in the 2MeSADP complex allows formation of an extensive ionic and polar interaction network with the phosphate groups of 2MeSADP. Connected by a disulphide bond to helix VII, the N terminus is moved towards the axis of the helical bundle as well; the position of R19, for example, shifts by ~6.6 Å with agonist bound. Compared with the structure of the antagonist-bound protease-activated receptor 1 (PAR1, also known as F2R, the receptor with the closest sequence homology and known structure to $P2Y_{12}R$)[19], the largest differences are observed in the straight versus bent conformation of helix V, as well as in the positions of the extracellular parts of helix VI, which in PAR1 adopts an intermediate conformation between the agonist- and antagonist-bound $P2Y_{12}R$ structures (Extended Data Fig. 3).

All ECLs have clear density and are resolved in the agonist-bound structure, whereas only ECL3 was resolved in the $P2Y_{12}R$–AZD1283 complex structure. The highly conserved $C97^{3.25}$–$C175^{ECL2}$ disulphide bond stabilizing the conformation of ECL2 is clearly observed in the agonist-bound structure, although it was not resolved with antagonist bound[5]. Formation of this disulphide bond in the $P2Y_{12}R$–2MeSADP complex requires an unwinding of the helical bulge structure in the extracellular portion of helix III that comprises ~4.6 residues instead of the ~3.6 residues in a regular α-helical turn. As a result, there is a Cα rotation of $C97^{3.25}$ along the helical path by over 60°, and a relocation of other residues at this region as compared with the AZD1283 complex

[1]CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Pudong, Shanghai 201203, China. [2]Molecular Recognition Section, Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [4]PharmaCenter Bonn, University of Bonn, Pharmaceutical Chemistry I, An der Immenburg 4, D-53121 Bonn, Germany. [5]Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Pudong, Shanghai 201203, China. [6]iHuman Institute, ShanghaiTech University, 99 Haike Road, Pudong, Shanghai 201203, China.
*These authors contributed equally to this work.

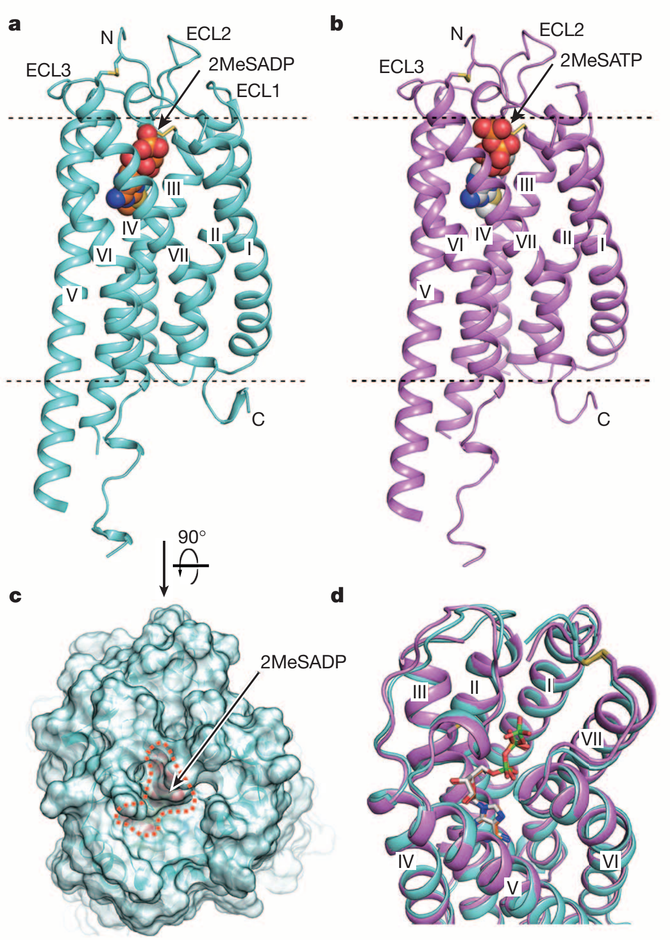


Figure 1 | Structure of the P2Y$_{12}$R–2MeSADP and P2Y$_{12}$R–2MeSATP complexes. **a**, Side view of P2Y$_{12}$R–2MeSADP complex structure. The receptor is coloured cyan and shown in cartoon representation. The ligand 2MeSADP is shown in sphere representation with orange carbons. The disulphide bonds are shown as yellow sticks, extracellular and intracellular boundaries are shown as dashed lines. **b**, Side view of the P2Y$_{12}$R–2MeSATP complex structure. The receptor is coloured violet and shown in cartoon representation. The ligand 2MeSATP is shown in sphere representation with grey carbons. **c**, Semi-transparent surface presentation of the receptor shows the lid formed by ECL2, ECL3 and N terminus on top of 2MeSADP. **d**, Comparison of P2Y$_{12}$R–2MeSADP with the P2Y$_{12}$R–2MeSATP complex. The ligands are shown as sticks, and the phosphorus in 2MeSATP is coloured in green to allow it to be distinguished from 2MeSADP.

(Extended Data Fig. 4). In addition to the structural changes of helix III, the extracellular tip of helix V is also shifted ~2 Å towards the helical bundle to satisfy formation of this disulphide bond (Fig. 2b).

With these substantial rearrangements of the helical bundle, ECLs and the N terminus, the agonist-bound structure appears to be much tighter compared with the antagonist-bound structure, with 2MeSADP almost completely enclosed in the receptor. A trend for slight contraction of the ligand binding pocket was observed for some α-group GPCRs in agonist-bound as compared to antagonist-bound structures[7–16], though these extracellular rearrangements are of unprecedented scale in the P2Y$_{12}$R. This very high plasticity of the extracellular region in the P2Y$_{12}$R may help to explain how rather bulky nucleotide ligands gain access into the binding cavity, which is completely covered by loops and N-terminal residues (Fig. 1c). A high plasticity in the ECL region was also proposed for another δ-group GPCR, PAR1 (ref. 19), but the current availability of only an antagonist-bound PAR1 structure precludes crystallographic description of those changes.

Interestingly, the intracellular conformational changes of P2Y$_{12}$R are less prominent than those at the extracellular side (Fig. 2b, c). In both P2Y$_{12}$R structures the intracellular half of the 7TM domain has a very similar conformation to that observed in the PAR1 structure (Extended Data Fig. 3c). Only minor changes in helices VI and VII were observed in the intracellular part of P2Y$_{12}$R–2MeSADP. They do not appear consistent, however, with large changes in helical positions observed in the intracellular region in active state agonist-bound A$_{2A}$AR, or β$_2$AR stabilized by G-protein[12,14]. It is therefore likely that the P2Y$_{12}$R–2MeSADP structure represents an agonist-bound inactive state with respect to the intracellular region, similar to the one observed in agonist-bound β$_1$AR and β$_2$AR without G-protein or a G-protein mimic stabilizing their active state[10,20].

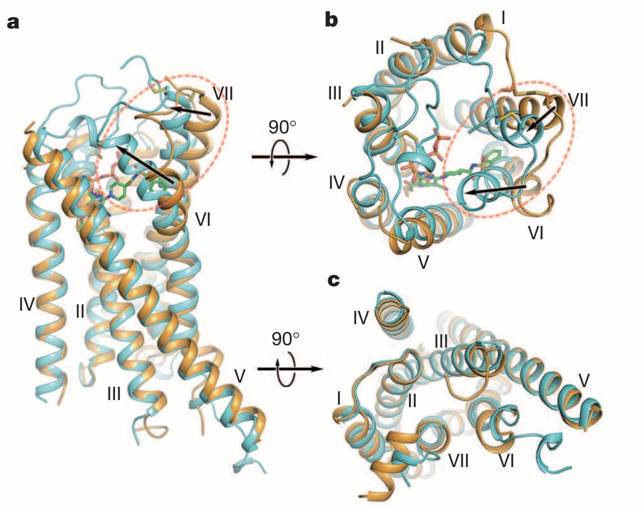


Figure 2 | Comparison of the P2Y$_{12}$R–2MeSADP (agonist) and P2Y$_{12}$R–AZD1283 (antagonist) complexes. **a**, The P2Y$_{12}$R–2MeSADP complex (receptor shown as cyan cartoon and ligand as sticks of orange carbons) and P2Y$_{12}$R–AZD1283 complex (receptor shown as light orange cartoon and ligand as sticks of green carbons) are shown in side view. Movement of the extracellular tips of helices VI and VII towards the centre of the 7TM domain is shown by arrows. The extracellular (**b**) and intracellular (**c**) views of the comparison are also shown.

The rearranged 2MeSADP-binding pocket consists of residues from helices III, IV, V, VI and VII as in the P2Y$_{12}$R–AZD1283 structure[5], but also extensively involves ECL2 and the N terminus. Both pockets described in the P2Y$_{12}$R–AZD1283 structure are still present, although contracted, especially pocket 1, due to the inward movement of helices VI and VII (Extended Data Fig. 5). In particular, the inward shift of helix VI in the agonist-bound structure shrinks pocket 1 substantially so that it would preclude AZD1283 binding. As a result, although both 2MeSADP and AZD1283 bind to the same pocket, their orientations are completely different, with only partial overlap between them (Fig. 3).

The adenine group of 2MeSADP occupies the same aromatic binding site as the nicotinate group in AZD1283, forming a similar π–π interaction with the Y105$^{3.33}$ side chain. The 2-thioether inserts into a hydrophobic pocket formed by F106$^{3.34}$, L155$^{4.56}$, S156$^{4.57}$ and N159$^{4.60}$, and serves as an anchor to maintain the adenine core and the ribose ring in an optimal orientation (Extended Data Fig. 5a). Thus, 2MeSADP binds with greater complementarity, which explains the higher affinity of this ligand compared with ADP. The 2-thioether and amino groups of adenine overlay with the ethyl ester and cyano substituents on the nicotinate group (Fig. 3b). The orientation of the ribose moiety corresponds to that of the methyl group on the nicotinate moiety. The ribose 2′ and 3′ hydroxyl groups interact with K179^{ECL2} and H187$^{5.36}$ and with T163^{ECL2} and K179^{ECL2}, respectively.

The interactions between the receptor and the diphosphate of 2MeSADP involve numerous hydrophilic and positively charged residues (Fig. 3c). As predicted on the basis of the sequence analysis and confirmed through mutagenesis[21–23], two essential cationic residues R256$^{6.55}$ and K280$^{7.35}$ (Extended Data Table 2) and an aromatic residue Y259$^{6.58}$ that is conserved

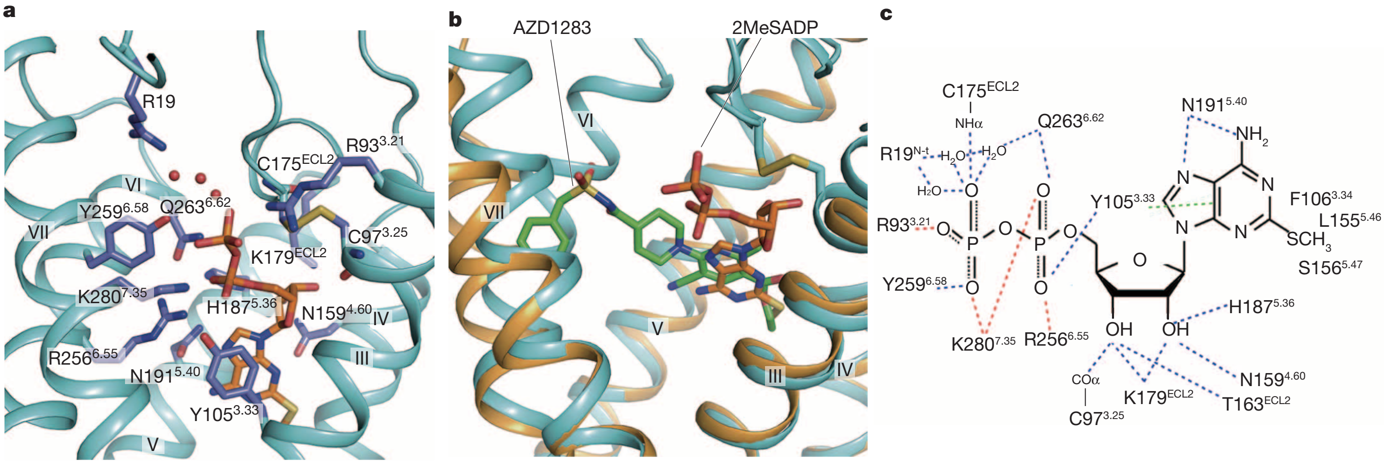


Figure 3 | $P2Y_{12}R$ ligand binding pocket for 2MeSADP. **a**, The receptor is shown in cyan cartoon representation. The ligand 2MeSADP (orange carbons) and receptor residues (slate carbons) involved in ligand binding are shown in stick representation. Other elements are coloured as follows: oxygen, red; nitrogen, dark blue; sulphur, yellow; phosphorus, orange. The water molecules interacting with 2MeSADP are shown as red spheres. **b**, Comparison of the 2MeSADP and AZD1283 binding poses in the overlaid $P2Y_{12}R$ complexes. Colour scheme as in Fig. 2. **c**, Summary of receptor interactions of 2MeSADP. Hydrogen bonds are displayed as blue dashed lines and the salt bridges as red dashed lines. The π–π interaction between 2MeSADP and $Y105^{3.33}$ is indicated as green dashed lines. The NHα and COα indicate the main chain amine and carbonyl groups of the corresponding residue.

in the $P2Y_{12}R$-like subfamily coordinate phosphate moieties. In addition to these residues, a residue not previously implicated in agonist binding, the third cationic residue $R93^{3.21}$, contacts the β-phosphate. Three water molecules also bridge the interaction between β-phosphate with a fourth cationic residue, R19 (N terminus). Some residues that are thought to participate in agonist binding, however, have no direct contact with the nucleotide ligands. The $R265^{ECL3}$ side chain, which was previously implicated in activation of the receptor[23], is positioned away from 2MeSADP. Interestingly, the conserved $K174^{ECL2}$, which was previously predicted to interact with 2MeSADP[22,24], does not form a direct contact with the agonist, but forms a salt bridge to $E273^{7.28}$ that apparently helps to stabilize the agonist-bound conformation.

Cysteine residues forming the conserved disulphide bond are also involved directly in the binding of 2MeSADP. The main chain carbonyl of $C97^{3.25}$ forms a hydrogen bond with the 3′ hydroxyl group of ribose, and the main chain NH group of $C175^{ECL2}$ interacts with the β-phosphate group. Disruption of this labile disulphide bond, for example, as possible in the complex with antagonist AZD1283, would also prevent such a bidentate coordination system of the nucleoside. The active metabolites of the thienopyridine drugs are also predicted to destabilize interactions of the $P2Y_{12}R$ with the nucleotide agonist by binding covalently to $C97^{3.25}$. Mutation of either of these cysteine residues to alanine greatly reduces agonist binding (Extended Data Table 2), although it is known that the $P2Y_{12}R$ does not require this disulphide bond for overall structural integrity, that is, in the AZD1283 complex.

The action of nucleoside 5′-triphosphate derivatives at $P2Y_{12}R$, that is, agonism versus antagonism, has long been unclear. Although ATP and 2MeSATP were thought to be antagonists under physiological conditions[25], the close conformational similarity between the 2MeSADP and 2MeSATP complexes suggests that ATP derivatives potentially qualify for similar signalling properties at $P2Y_{12}R$ as ADP derivatives, which is also supported by recent pharmacological data[3]. Moreover, AR-C66096, a non-cleavable triphosphate mimic that could be docked into the $P2Y_{12}R$ in an identical binding mode to 2MeSADP and 2MeSATP, displays characteristics of a partial agonist inhibiting cAMP production in CHO cells overexpressing $P2Y_{12}R$ (Extended Data Fig. 6). Of course, we emphasize that partial agonist activity of ATP derivatives may be expression-level dependent[26].

The 2MeSADP binding cavity appears to accommodate a number of other nucleotide ligands that mimic a triphosphate chain, including

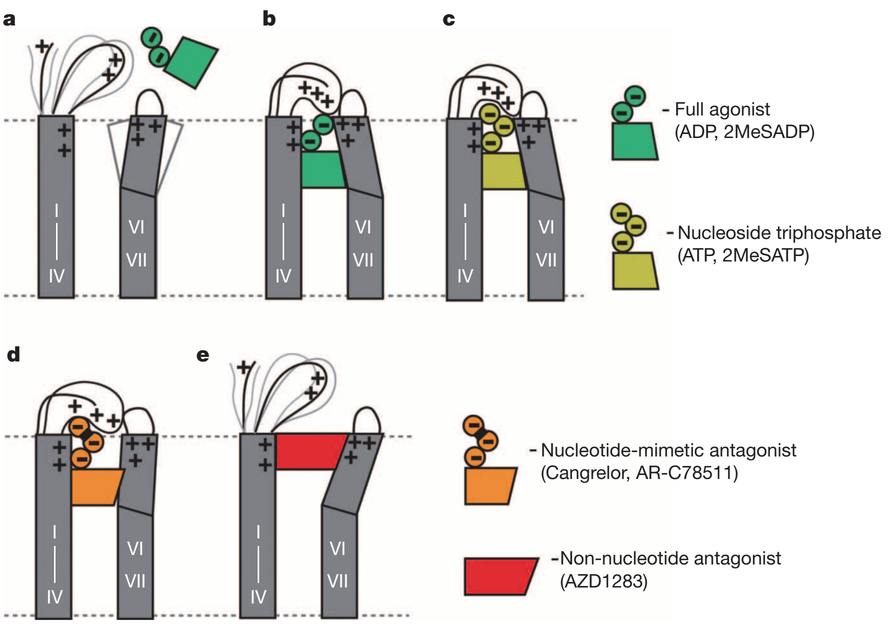


Figure 4 | Schematic illustration of conformational changes in $P2Y_{12}R$ extracellular region. **a**, Unliganded (apo) state of $P2Y_{12}R$ with open entrance to the pocket and partially disordered lid. A number of partially uncompensated positive electrostatic charges among the side chains in the pocket ($R19^{Nterm}$, $K80^{2.60}$, $R93^{3.21}$, $K173^{ECL2}$, $K174^{ECL2}$, $R256^{6.55}$ and $K280^{7.35}$) disfavour formation of the stable closed state. **b**, The closed state is stabilized by binding of nucleotide agonist (for example, ADP, 2MeSADP). **c**, A similar conformation with 'lid' closure occurs in 2MeSATP structure and for various docked N^6 unsubstituted nucleoside triphosphate and triphosphate-mimetic ligands. **d**, A helical reorganization is proposed for some N^6 substituted nucleoside triphosphate and triphosphate mimetic ligands, especially with bulky N^6 substituents. **e**, Binding of non-nucleotide antagonist AZD1283 blocks inward movement of helices VI and VII and prevents 'lid' closure.

AR-C67085 and cangrelor[27] (Extended Data Figs 7 and 8). However, the nucleoside antagonist ticagrelor does not dock with a similar conformation to 2MeSADP in the $P2Y_{12}R$ because of the bulky N^6 substituent, unless a helical rearrangement occurs, particularly in helix VI. This docking observation suggests that reduction of agonist efficacy of these ligands is likely facilitated by N^6 substitutions in the nucleoside scaffold[27]. Consistently, we have shown that non-nucleotides AZD1283 and ticagrelor behave as competitive $P2Y_{12}R$ antagonists. Thus, we predict that the ligand-bound receptor conformations and consequently pharmacology of nucleotide-mimetic and non-nucleotide antagonists will be divergent.

These structural and pharmacological findings provide unexpected insights into the mechanism of agonist and antagonist interactions with $P2Y_{12}R$, as schematically illustrated in Fig. 4. As the agonist-bound pocket is sealed by a 'lid' formed by unusually cationic ECLs and the N terminus, agonist access to the binding pocket of apo $P2Y_{12}R$ would require plasticity of this extracellular region. Closing of the pocket in the absence of charged phosphates would also be disfavoured by electrostatic repulsion, with $\sim$7 arginine and lysine side chains pointing towards the pocket. Binding of a nucleotide agonist like 2MeSADP involves stabilization of an inward position of helices VI and VII and formation of the lid, which is stabilized by numerous electrostatic interactions with the agonist phosphate groups. The additional phosphate group of 2MeSATP is also accommodated by a similar conformation of the lid, although distinct interactions may still affect the 2MeSATP binding and signalling profile. In stark contrast, an open extracellular side conformation is defined by binding of non-nucleotide antagonists like AZD1283, which keeps helices VI and VII away from the pocket and destabilizes the lid.

In conclusion, $P2Y_{12}R$ is the first receptor to demonstrate striking rearrangements in the extracellular regions, characterized by open and closed access to the ligand binding pocket. Whether such high plasticity of the binding pocket has evolved to enable optimal and specific recognition of negatively charged nucleotide ligands of the P2YR family, or is a more general feature of the δ-group of class A GPCRs is currently unknown. Additional structures will help us to understand similarities and differences in ligand binding and receptor activation of the δ-group receptors.

METHODS SUMMARY

The human $P2Y_{12}R$–BRIL fusion protein was purified by metal affinity chromatography and crystalized in lipidic cubic phase (LCP). The diffraction data were collected from 17 crystals ($P2Y_{12}R$–2MeSADP) and 6 crystals ($P2Y_{12}R$–2MeSATP) on the BL41XU beamline at the SPring8, Japan, using a 10 μm minibeam. The structure was solved by molecular replacement and refined in Refmac5 and Buster.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 January; accepted 28 March 2014.

1. Chang, H. *et al.* Modified diadenosine tetraphosphates with dual specificity for $P2Y_1$ and $P2Y_{12}$ are potent antagonists of ADP-induced platelet activation. *J. Thromb. Haemost.* **10,** 2573–2580 (2012).
2. Chang, H. *et al.* Agonist and antagonist effects of diadenosine tetraphosphate, a platelet dense granule constituent, on platelet $P2Y_1$, $P2Y_{12}$ and $P2X_1$ receptors. *Thromb. Res.* **125,** 159–165 (2010).
3. Schmidt, P. *et al.* Identification of determinants required for agonistic and inverse agonistic ligand properties at the ADP receptor $P2Y_{12}$. *Mol. Pharmacol.* **83,** 256–266 (2013).
4. Srinivasan, S. *et al.* The $P2Y_{12}$ antagonists, 2-methylthioadenosine 5′-monophosphate triethylammonium salt and cangrelor (ARC69931MX), can inhibit human platelet aggregation through a G_i-independent increase in cAMP levels. *J. Biol. Chem.* **284,** 16108–16117 (2009).
5. Zhang, K. *et al.* Structure of the human $P2Y_{12}$ receptor in complex with an antithrombotic drug. *Nature* http://dx.doi.org/101038/nature13083 (this issue).
6. Zhou, X. E., Melcher, K. & Xu, H. E. Structure and activation of rhodopsin. *Acta Pharmacol. Sin.* **33,** 291–299 (2012).
7. Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289,** 739–745 (2000).
8. Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455,** 497–502 (2008).
9. Warne, T. *et al.* Structure of a β_1-adrenergic G-protein-coupled receptor. *Nature* **454,** 486–491 (2008).
10. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β_1-adrenergic receptor. *Nature* **469,** 241–244 (2011).
11. Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **318,** 1258–1265 (2007).
12. Rasmussen, S. G. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477,** 549–555 (2011).
13. Jaakola, V. P. *et al.* The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322,** 1211–1217 (2008).
14. Xu, F. *et al.* Structure of an agonist-bound human A2A adenosine receptor. *Science* **332,** 322–327 (2011).
15. Haga, K. *et al.* Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482,** 547–551 (2012).
16. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504,** 101–106 (2013).
17. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63,** 1256–1272 (2003).
18. Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* Vol. 25 (ed. C. Sealfon Stuart) 366–428 (Academic Press, 1995).
19. Zhang, C. *et al.* High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492,** 387–392 (2012).
20. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist–β_2 adrenoceptor complex. *Nature* **469,** 236–240 (2011).
21. Hoffmann, K., Sixel, U., Di Pasquale, F. & von Kugelgen, I. Involvement of basic amino acid residues in transmembrane regions 6 and 7 in agonist and antagonist recognition of the human platelet $P2Y_{12}$-receptor. *Biochem. Pharmacol.* **76,** 1201–1213 (2008).
22. Cattaneo, M. The platelet $P2Y_{12}$ receptor for adenosine diphosphate: congenital and drug-induced defects. *Blood* **117,** 2102–2112 (2011).
23. Ignatovica, V., Megnis, K., Lapins, M., Schioth, H. B. & Klovins, J. Identification and analysis of functionally important amino acids in human purinergic 12 receptor using a *Saccharomyces cerevisiae* expression system. *FEBS J.* **279,** 180–191 (2012).
24. Daly, M. E. *et al.* Identification and characterization of a novel $P2Y_{12}$ variant in a patient diagnosed with type 1 von Willebrand disease in the European MCMDM-1VWD study. *Blood* **113,** 4110–4113 (2009).
25. Kauffenstein, G., Hechler, B., Cazenave, J. P. & Gachet, C. Adenine triphosphate nucleotides are antagonists at the P2Y receptor. *J. Thromb. Haemost.* **2,** 1980–1988 (2004).
26. Fujioka, M. & Omori, N. Subtleties in GPCR drug discovery: a medicinal chemistry perspective. *Drug Discov. Today* **17,** 1133–1138 (2012).
27. Ding, Z., Kim, S. & Kunapuli, S. P. Identification of a potent inverse agonist at a constitutively active mutant of human $P2Y_{12}$ receptor. *Mol. Pharmacol.* **69,** 338–345 (2006).

Acknowledgements This work was supported by the National Basic Research Program of China grants 2012CB910400, 2012CB518000 and 2014CB910400 (B.W., Q.Z.), the National Institutes of Health grants R01 AI100604 (B.W., Q.Z.) and U54 GM094618 (V.C., V.K., R.C.S.; Target GPCR-87), the National Science Foundation of China grants 31370729 and National Science and Technology Major Project 2013ZX09507001 and 2012ZX09301001 (B.W., Q.Z.), National Institutes of Health NIDDK Intramural Research Program (K.A.J.) and the National Natural Science Foundation of China 91313000 (H.J.). The authors thank S. Nylander, E. Kiselev and S. Moss for scientific feedback on the manuscript, A. Walker for assistance with manuscript preparation and K. Kadyshevskaya for help with figure preparation. The synchrotron radiation experiments were performed at the BL41XU of SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal no. 2013B1049). We thank the beamline staff members of the BL41XU for help with X-ray data collection.

Author Contributions J.Z. optimized the construct, expressed and purified human $P2Y_{12}R$–BRIL for crystallization, developed the purification procedure, performed crystallization trials and optimized crystallization conditions. K.Z. helped with construct and crystal optimization and collected diffraction data. Z.-G.G. designed, performed and analysed ligand binding and competition assays of wild-type and mutant $P2Y_{12}R$. S.P. performed and analysed docking assays. D.Z. helped in expression and purification. G.W.H. solved and refined the structure. T.L. helped the expression for crystallization trials. L.M. helped the expression for the activity assays. W.Z. developed the initial expression and purification protocol for $P2Y_{12}R$. C.E.M. helped to design and analyse pharmacological experiments and wrote the manuscript. H.Y. helped to design and analyse docking assays. H.J. oversaw design and validation of $P2Y_{12}R$ models. V.C. helped to design and optimize LCP crystallization trials and processed crystallographic data and wrote the manuscript. V.K. performed and analysed molecular modelling simulations and wrote the manuscript. K.A.J. oversaw, designed and analysed ligand binding assays and docking and wrote the manuscript. R.C.S. oversaw expression, purification and crystallization, and structure analysis and interpretation of $P2Y_{12}R$. B.W. and Q.Z. initiated the project, planned and analysed experiments, supervised the research and wrote the manuscript.

Author Information Atomic coordinates and structure factors for the $P2Y_{12}R$–2MeSADP and $P2Y_{12}R$–2MeSATP structures have been deposited in the Protein Data Bank with identification codes 4PXZ and 4PY0, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.W. (beiliwu@simm.ac.cn) or Q.Z. (zhaoq@simm.ac.cn).

METHODS

Purification of $P2Y_{12}R$–BRIL protein and crystallization in lipidic cubic phase. $P2Y_{12}R$–BRIL construction, expression and membrane preparation were performed using the same procedure as described in the companion manuscript[5]. In brief, the human $P2Y_{12}R$ was subcloned into a modified pFastBac1 vector, with thermostabilized BRIL (Protein Data Bank accession code 1M6T) inserted at ICL3 (T223-R224) and a $D294^{7.49}N$ mutation. The fusion protein was expressed using Bac-to-Bac Baculovirus Expression System (Invitrogen) in *Spodoptera frugiperda* (*Sf*9) cells for 48 h and membrane was washed repeatedly using hypotonic buffer with low and high salt.

Before solubilization, purified membranes were incubated with 20 μM corresponding ligand (2MeSADP or 2MeSATP obtained from Tocris) in the presence of 2 mg ml^{-1} iodoacetamide, and EDTA-free protease inhibitor cocktail (Roche) for 30 min. $P2Y_{12}R$–BRIL was extracted from the membrane by adding *n*-dodecyl-β-D-maltopyranoside (DDM, Affymetrix) and cholesteryl hemisuccinate (CHS, Sigma) to the membrane solution to a final concentration of 0.5% (w/v) and 0.1% (w/v), respectively, and stirring was continued at 4 °C for 2.5 h. The supernatant was isolated by centrifugation at 160,000*g* for 30 min, followed by incubation in TALON IMAC resin (Clontech) at 4 °C, overnight. The resin was washed with twenty column volumes of 50 mM HEPES, pH 7.5, 1 M NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS, and 30 mM imidazole. The protein was eluted with 5 column volumes of 50 mM HEPES, pH 7.5, 1 M NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS, 270 mM imidazole and 50 μM corresponding ligand. After removing imidazole by using a PD MiniTrap G-25 column (GE Healthcare), ligand concentration was increased to 2 mM. The protein was then treated overnight with His-tagged PreScission protease (20 μg per 500 ml of expressed material) and His-tagged PNGase F (20 μg per 500 ml of expressed material) to remove the carboxy-terminal His-tag and deglycosylate the receptor. PreScission protease, PNGase F and the cleaved 10× His-tag were removed from the sample by passing the sample over Ni-NTA superflow resin (Qiagen). The protein was then concentrated to 30–40 mg ml^{-1} with a 100 kDa molecular mass cut-off concentrator (Millipore).

Protein sample was reconstituted into lipidic cubic phase (LCP) by mixing 40% of ~30 mg ml^{-1} protein with 60% lipid (10% (w/w) cholesterol, 90% (w/w) monoolein)[28]. Crystallization trials were performed using a syringe lipid mixer as previously described[29]. The protein–lipid mixture was dispensed in 40 nl drops onto glass sandwich plates (Shanghai FAstal BioTech) and overlaid with 800 nl precipitant solution using a Mosquito LCP robot (TTP LabTech). For $P2Y_{12}R$–2MeSADP–BRIL complex, crystals appeared after 1 week in 0.30–0.45 M ammonium acetate, 0.1 M sodium citrate, pH 5.0, 30–40% PEG400, 3% v/v 1-propanol and 500 μM 2MeSADP and reached their full size ($80 \times 50 \times 5\ \mu m^3$) within 2 weeks (Extended Data Fig. 1). For $P2Y_{12}R$–2MeSATP–BRIL complex, crystals were obtained from precipitant conditions containing 0.15–0.20 M ammonium tartrate, 0.1 M sodium citrate, pH 6.0, 35–40% PEG400, 4% v/v MPD and 500 μM 2MeSATP, reaching their full size ($30 \times 30 \times 5\ \mu m^3$) within 10 days. Crystals were collected directly from LCP using 50–150 μm micromounts (M2-L19-50/150, MiTeGen) and flash frozen in liquid nitrogen.

Data collection, structure solution and refinement. X-ray diffraction data were collected at the SPring-8 beam line 41XU, Hyogo, Japan, using a Rayonix MX225HE detector (X-ray wavelength 1.0000 Å). The crystals were exposed with a 10 μm minibeam for 1 s and 1° oscillation per frame, and a rastering system was used to find the best diffracting parts of single crystals[30]. Most crystals of $P2Y_{12}R$ in complex with 2MeSADP diffracted to 3.0–2.5 Å resolution. XDS[31] was used for integrating and scaling data from 17 best-diffracting crystals for $P2Y_{12}R$–2MeSADP complexes and 6 crystals for $P2Y_{12}R$–2MeSATP complexes. The $P2Y_{12}R$-2MeSADP complex was solved by molecular replacement with Phaser[32] using the receptor portion of PAR1 (PDB: 3VW7), converted to polyalanines, and BRIL (PDB: 1M6T) as initial models and refined in Refmac5[33] and Buster[34]. The $P2Y_{12}R$–2MeSATP structure was solved using $P2Y_{12}R$ in complex with 2MeSADP and BRIL (PDB: 1M6T) as starting models and refined under the same procedure.

Ligand-binding assays. Wild-type (WT) and mutant human $P2Y_{12}R$ plasmids with single amino acid substitutions were cloned into pcDNA3.0 vector and transfected into COS7 cells using Lipofectamine 2000 (Life Technologies). Cells were collected 48 h after transfection. After collection, cells were homogenized for 15 s and then centrifuged for 10 min at 1,000*g*. The suspension was re-centrifuged at 20,000*g* for 60 min. The resulting pellet was resuspended, homogenized, split into aliquots and maintained at –80 °C in a freezer until use. Protein concentrations were measured using Bio-Rad protein assay reagents.

For saturation experiments, 50 μl [^{3}H]2MeSADP (3.5 Ci $mmol^{-1}$, from 0.4 to 46 nM, Moravek) was incubated with 100 μl WT and mutant $P2Y_{12}R$ membrane preparations (5 μg per tube) in a total assay volume of 200 μl Tris·HCl buffer containing 10 mM $MgCl_2$. AZD1283 (10 μM) was used to determine the non-specific binding. For displacement experiments, increasing concentrations of AZD1283 were incubated with WT or mutant membrane preparations (5–10 μg) and [^{3}H]2MeSADP (10 nM) at 25 °C for 30 min. ADP, 2MeSADP and 2MeSATP were obtained from Sigma. AR-C66096 (>98%) was obtained from Tocris. The reaction was terminated by filtration with a 24-channnel Brandel cell harvester (Brandel) and washed twice with 5 ml cold Tris·HCl buffer containing 10 mM $MgCl_2$. Radioactivity was measured using a scintillation counter (Tri-Carb 2810TR). Data were analysed using Prism 6 (GraphPad).

Cyclic AMP (cAMP) accumulation assay. CHO cells stably expressing the human $P2Y_{12}R$ were plated in 96-well plates in 0.1 ml medium. After overnight incubation, the medium was removed, and cells were washed three times with 0.1 ml Hank's buffer, pH 7.4. Cells were then treated with the antagonists in the presence of rolipram (10 μM) for 20 min before the addition of agonists. Nucleotide derivatives were incubated with cells for 15 min followed by the addition of forskolin (10 μM) and incubation for another 15 min. The reaction was terminated by removing the supernatant, and cells were lysed upon the addition of 100 μl of 0.3% Tween-20. For determination of cAMP production, the ALPHA Screen cAMP assay kit (Perkin-Elmer) was used following the instructions provided with the kit.

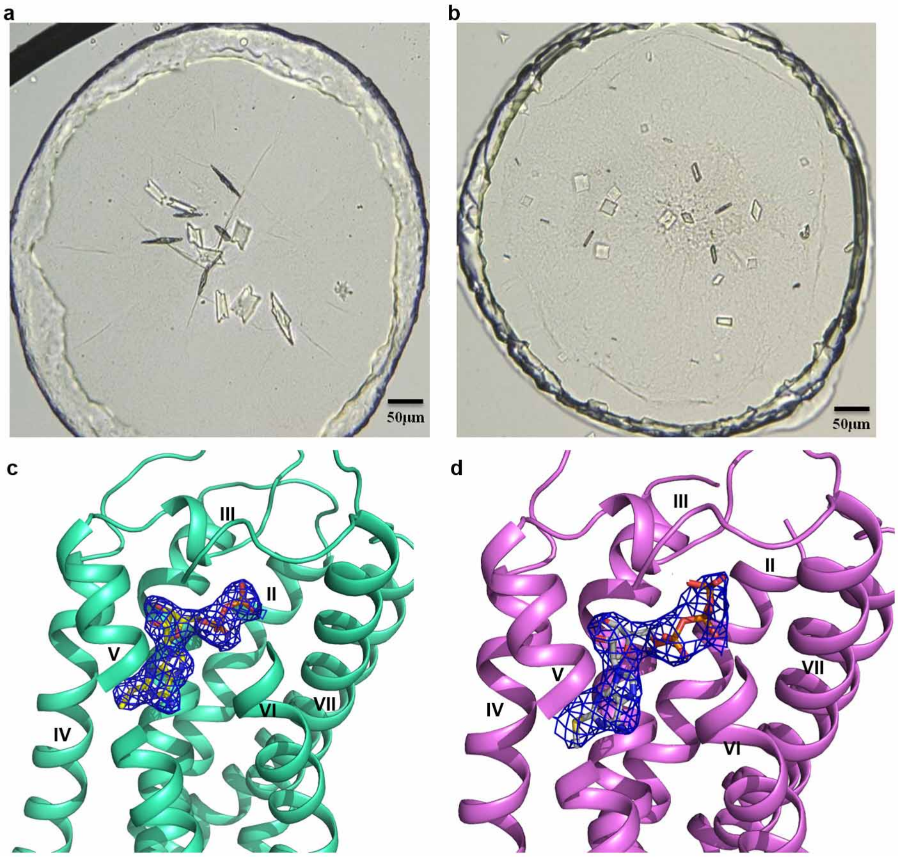
Docking. The $P2Y_{12}R$–2MeSADP structure was prepared using the Protein Preparation Wizard tool implemented in the Schrödinger suite[35], adding all the hydrogen atoms and the missing side chains of residues whose backbone coordinates were observed in the structure. The BRIL portion was removed. The orientation of polar hydrogens was optimized, the protein protonation states were adjusted and the overall structure was minimized with harmonic restraints on the heavy atoms, to remove strain. Then, all the hetero groups and water molecules were deleted.

The SiteMap tool of the Schrödinger suite was used to identify potential binding sites in the structure. Molecular docking of selected compounds (ADP, 2MeSADP, ATP, 2MeSATP, AR-C67085, AR-C66096, cangrelor and ticagrelor) at the $P2Y_{12}R$ structure was performed by means of the Glide package from the Schrödinger suite. In particular, a Glide Grid was centred on the centroid of residues located within 6 Å from the previously identified cavity. The Glide Grid was built using an inner box (ligand diameter midpoint box) of 14 Å × 14 Å × 14 Å and an outer box that extended 10 Å in each direction from the inner one (so that ligands up to 20 Å could be docked). Docking of ligands was performed in the rigid binding site using the XP (extra precision) procedure. The top scoring docking conformations for each ligand were subjected to visual inspection and analysis of protein–ligand interactions to select the final binding conformations in agreement with the experimental data.

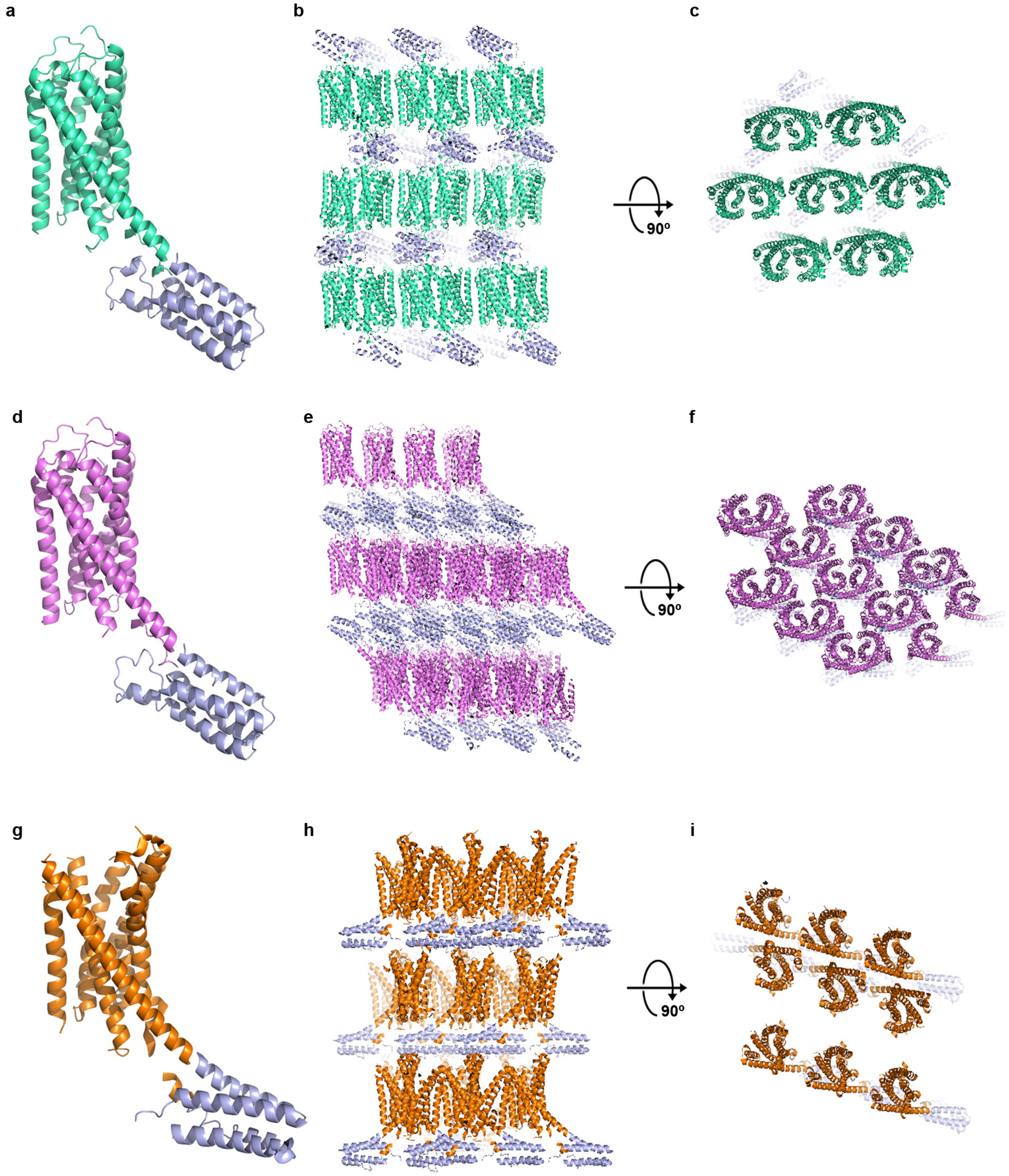
DNA sequence of the crystallization construct. The sequence was: ATGAAGACGATCATCGCCCTGAGCTACATCTTCTGCCTGGTGTTCGCCGACTACAAGGACGATGATGACGGCGCGCCGCAAGCCGTGGACAACCTCACATCAGCCCCTGGCAACACCTCCCTCTGTACCCGCGACTACAAGATCACACAAGTTCTCTTCCCCCTCCTCTACACAGTGTTGTTCTTCGTCGGCCTCATCACCAACGGATTGGCTATGCGTATCTTCTTCCAGATCCGCTCCAAGTCTAACTTCATCATCTTCCTGAAGAACACTGTGATCTCGGACCTGCTCATGATCCTCACATTCCCATTCAAGATCCTGTCAGATGCCAAGCTCGGTACTGGCCCGTTGCGTACATTCGTCTGCCAGGTTACCTCTGTGATCTTCTACTTCACTATGTACATCAGCATCTCATTCCTGGGTCTCATCACCATCGACAGGTACCAAAAGACCACTAGACCCTTCAAGACTAGCAACCCTAAGAACTTGCTGGGCGCTAAGATCCTGAGCGTGGTCATCTGGGCCTTCATGTTCCTCTTGTCACTGCCCAACATGATCCTCACCAACAGGCAGCCTAGAGATAAGAACGTGAAGAAGTGTTCATTCCTCAAGTCGGAGTTCGGATTGGTTTGGCACGAAATCGTGAACTACATCTGCCAAGTCATCTTCTGGATCAACTTCCTGATCGTTATCGTGTGTTACACATTGATCACCAAGGAGCTCTACAGGTCCTACGTCCGTACTGCTGATCTGGAAGACAATTGGGAAACTCTGAACGACAATCTCAAGGTGATCGAGAAGGCTGACAATGCTGCACAAGTCAAAGACGCTCTGACCAAGATGAGGGCAGCAGCCCTGGACGCTCAGAAGGCCACTCCACCTAAGCTCGAGGACAAGAGCCCAGATAGCCCTGAAATGAAAGACTTTCGGCATGGATTCGACATTCTGGTGGGACAGATTGATGATGCACTCAAGCTGGCCAATGAAGGGAAAGTCAAGGAAGCACAAGCAGCCGCTGAGCAGCTGAAGACCACCCGGAATGCATACATTCAGAAGTACCTGCGCGGAGTCGGCAAGGTTCCTAGGAAGAAGGTCAACGTTAAGGTGTTCATCATCATCGCTGTCTTCTTCATCTGCTTCGTTCCATTCCACTTCGCCCGTATCCCGTACACTTTGTCCCAAACACGCGACGTGTTCGATTGTACCGCTGAGAACACTCTGTTCTACGTCAAGGAATCCACATTGTGGCTGACCTCTCTGAACGCCTGCCTCAACCCATTCATCTACTTCTTCCTCTGTAAGTCTTTCCGCAACTCGTTGATCTCCATGCTGAAGTGCCCTAACTCTGCTACCAGCCTGTCCCAAGATAACAGAAAGAAGGAGCAGGACGGAGGCGACCCGAACGAGGAAACCCCGATGGGCCGGCCTCTGGAAGTTCTGTTCCAGGGGCCCCATCATCATCATCATCATCATCATCATCATTAG.

28. Zhao, Q. & Wu, B. L. Ice breaking in GPCR structural biology. *Acta Pharmacol. Sin.* **33,** 324–334 (2012).
29. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4,** 706–731 (2009).
30. Cherezov, V. *et al.* Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 μm

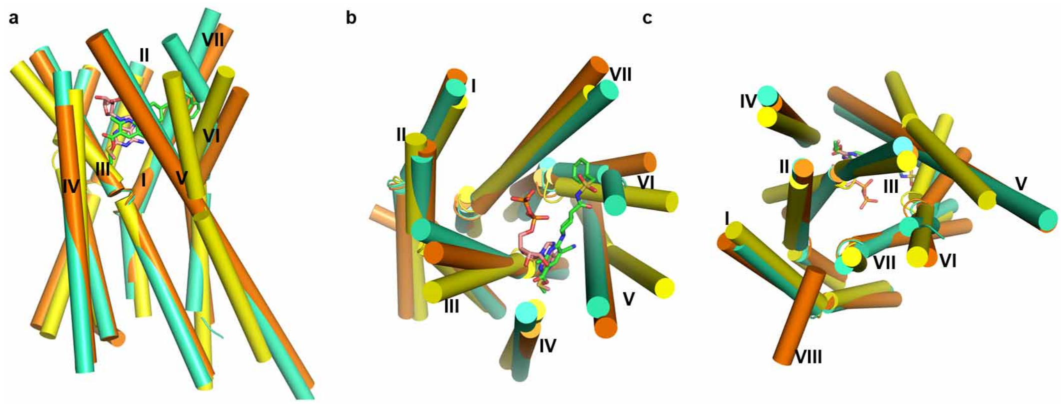
size X-ray synchrotron beam. *J. R. Soc. Interface* **6** (Suppl. 5), S587–S597 (2009).
31. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
33. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60,** 2184–2195 (2004).
34. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68,** 368–380 (2012).
35. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27,** 221–234 (2013).



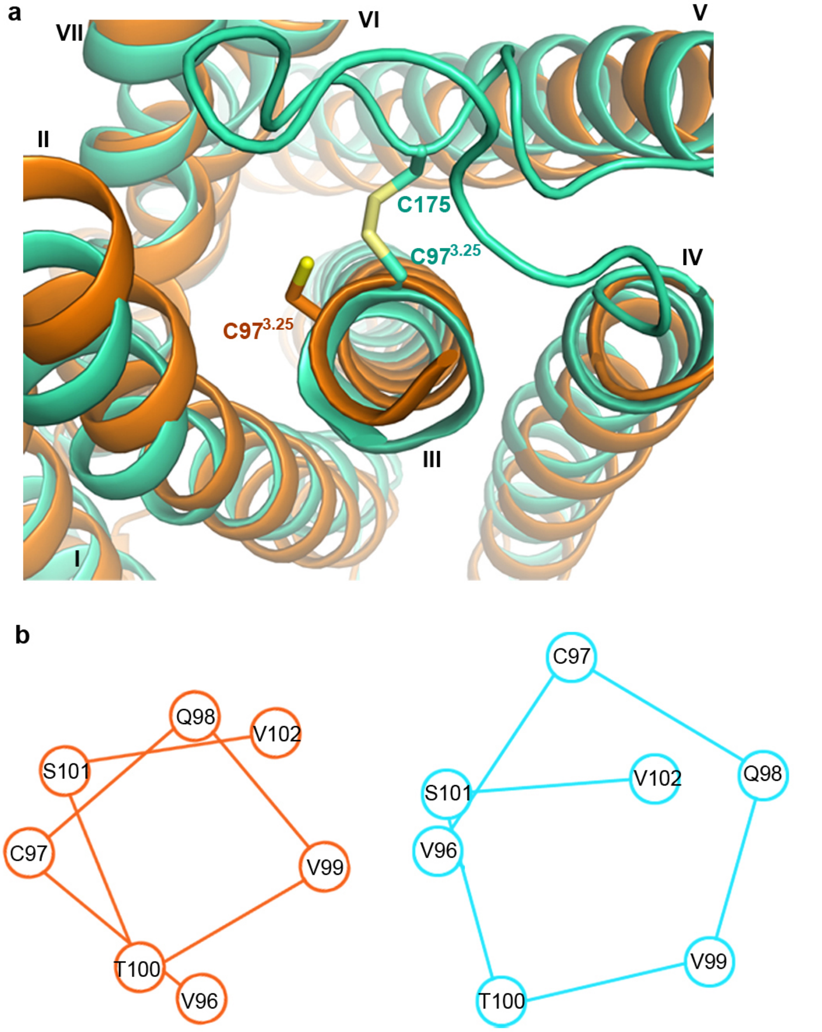
Extended Data Figure 1 | Crystals and electron density of nucleotides for P2Y$_{12}$R–2MeSADP and P2Y$_{12}$R–2MeSATP complexes. **a**, Crystals of the P2Y$_{12}$R–2MeSADP complex. The size of the crystals is roughly $80 \times 50 \times 5$ μm. **b**, Crystals of the P2Y$_{12}$R–2MeSATP complex. The size of the crystals is roughly $30 \times 30 \times 5$μm. **c**, The 2mF_o–DF_c map for the 2MeSADP contoured at 1σ. **d**, The 2mF_o–DF_c map for the P2Y$_{12}$R–2MeSATP contoured at 1σ. The relatively high B-factor of the γ-phosphate group (98 Å^2) compared with β-phosphate and surrounding protein atoms (~75 Å^2), and the propensity of 2MeSATP to hydrolyse to 2MeSADP suggest partial occupancy for the γ-phosphate group. However, given the differences in crystal forms and packing, as well as the clear density of the γ-phosphate group, the P2Y$_{12}$R–2MeSATP complex structure should provide relevant information about 2MeSATP binding.



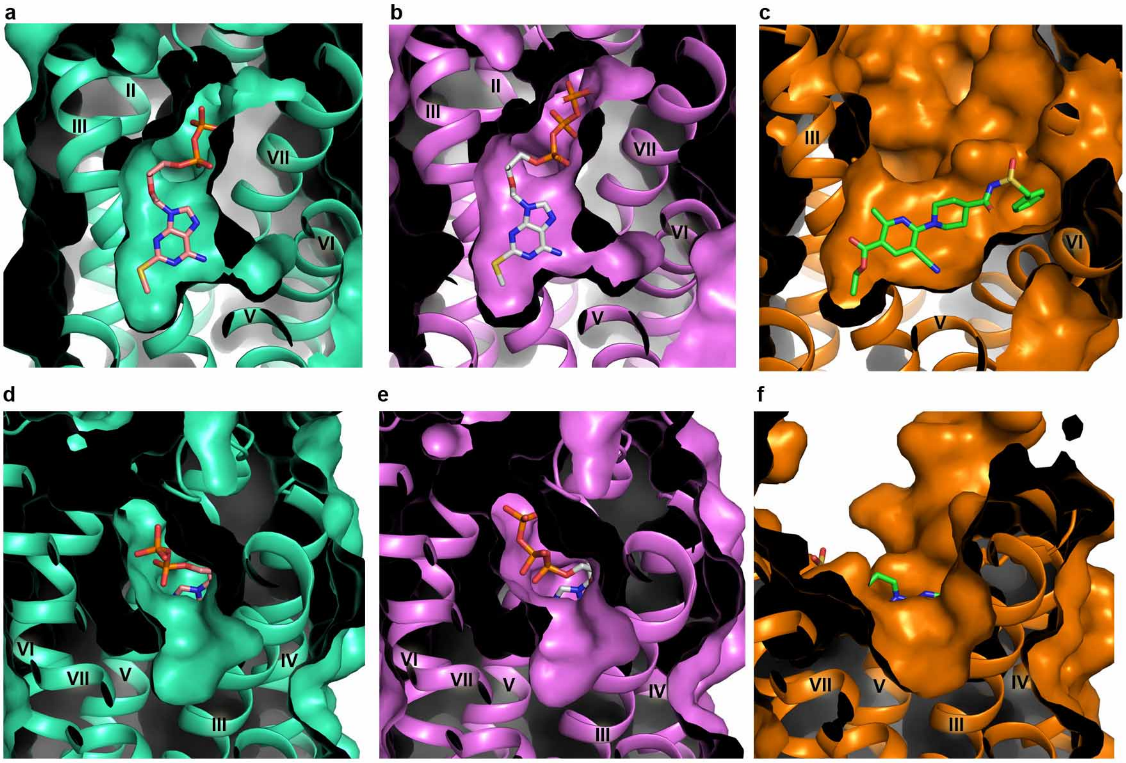
Extended Data Figure 2 | Crystal packing of P2Y$_{12}$R–2MeSADP, P2Y$_{12}$R–2MeSATP and P2Y$_{12}$R–AZD1283 complexes. **a**, Overall structure of the P2Y$_{12}$R–2MeSADP complex, P2Y$_{12}$R and BRIL are shown in cyan and blue, respectively. **b**, **c**, Crystal packing of P2Y$_{12}$R–2MeSADP complex shown in two different views. **d**, Overall structure of the P2Y$_{12}$R–2MeSATP complex, P2Y$_{12}$R is shown in pink. **e**, **f**, Crystal packing of P2Y$_{12}$R–2MeSATP complex in two different views. **g**, Overall structure of the P2Y$_{12}$R–AZD1283 complex, P2Y$_{12}$R is shown in orange. **h**, **i**, Crystal packing of P2Y$_{12}$R–AZD1283 complex shown in two different views.



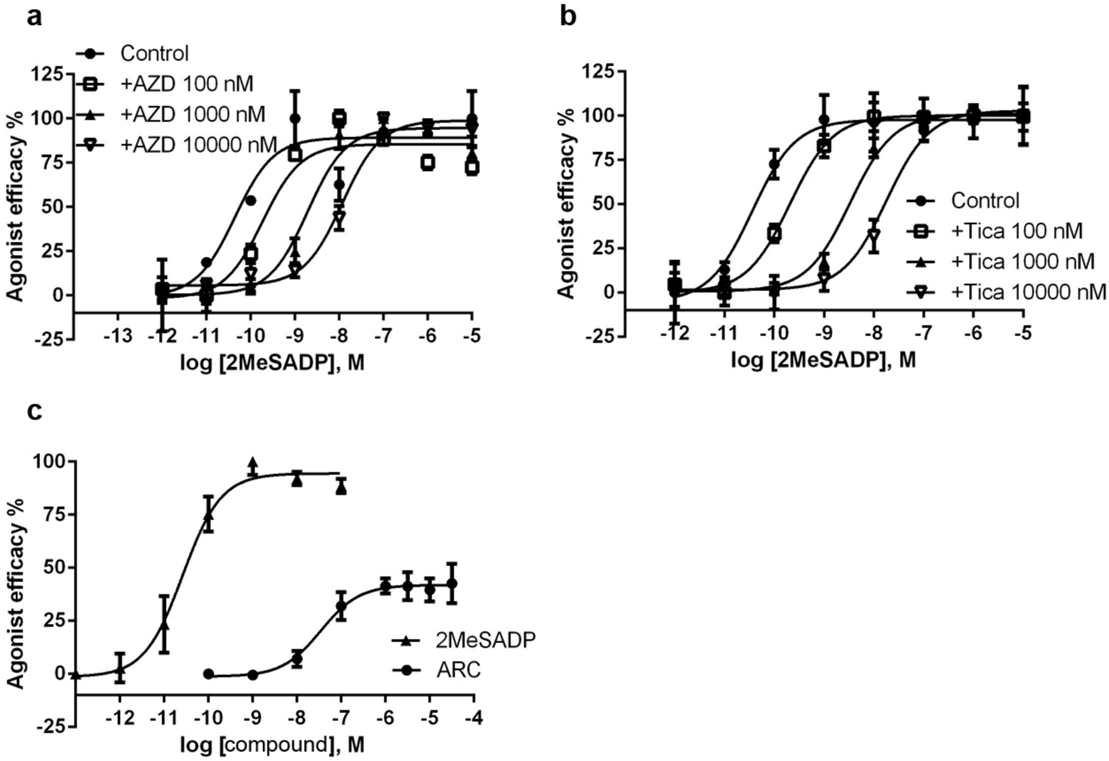
Extended Data Figure 3 | Comparison of antagonist- (orange) and agonist- (green/cyan) bound P2Y$_{12}$R structures with the PAR1 structure (yellow). **a**, Side view of the three structures. The receptor structures are shown as cylindrical helices and AZD1283 and 2MeSADP are shown as sticks with green carbons and wheat carbons, respectively. **b**, Comparison view from the extracellular side. **c**, Comparison view from the intracellular side.



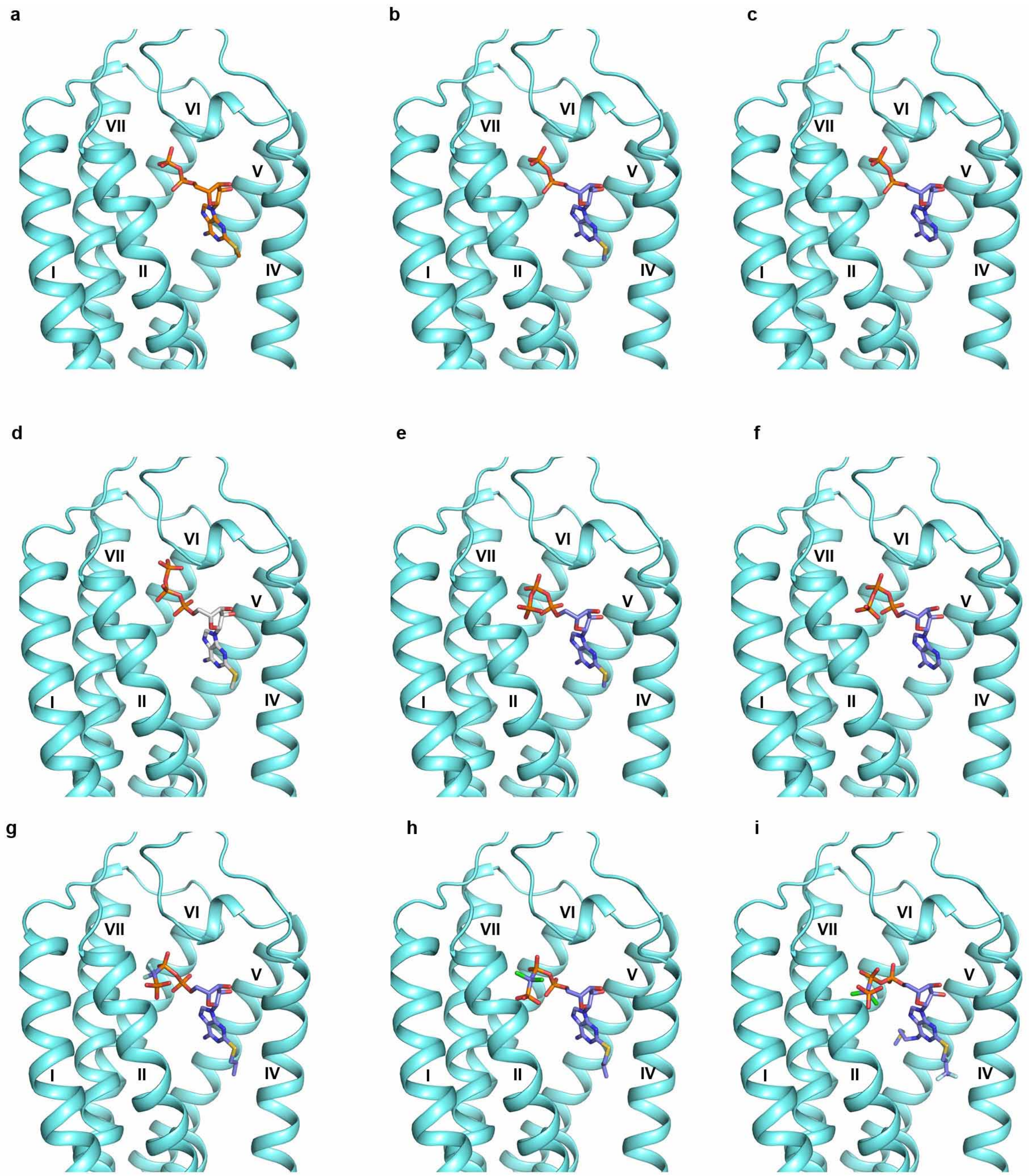
Extended Data Figure 4 | The distortion of helix III by the disulphide bond. **a**, Comparison of P2Y$_{12}$R–AZD1283 (orange) and P2Y$_{12}$R–2MeSADP (green/cyan). **b**, Corresponding positions of residues around C97$^{3.25}$ in P2Y$_{12}$R–AZD1283 (orange) and P2Y$_{12}$R–2MeSADP (green/cyan).



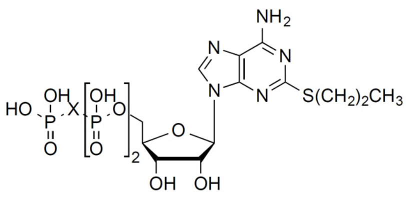
Extended Data Figure 5 | Comparison of pocket 1 (a–c) and pocket 2 (d–f) of P2Y$_{12}$R structures with different ligands. **a**, **d**, The P2Y$_{12}$R–2MeSADP structure. **b**, **e**, The P2Y$_{12}$R–2MeSATP structure. **c**, **f**, The P2Y$_{12}$R–AZD1283 structure. The 2MeSADP, 2MeSATP and AZD1283 ligands are shown in sticks with wheat, grey and green carbons, respectively.



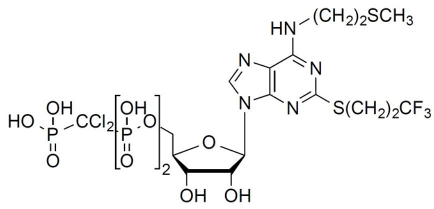
Extended Data Figure 6 | Functional properties of different ligands at P2Y$_{12}$R. Data (mean ± s.e.m.) were determined in triplicate. **a**, Parallel right shifts induced by antagonist AZD1283 (AZD) of the activation curves by agonist 2MeSADP in inhibition of cAMP production in P2Y$_{12}$R expressing CHO cells. **b**, Parallel right shifts induced by antagonist ticagrelor (Tica) of the activation curves by agonist 2MeSADP in inhibition of cAMP production in P2Y$_{12}$R expressing CHO cells. The pK_b values of AZD and Tica are 8.17 ± 0.45 and 7.70 ± 0.18, respectively. **c**, Partial agonist effects of AR-C66096 (ARC) in inhibition of cAMP production in P2Y$_{12}$R expressing CHO cells. The half-maximum effective concentration (EC$_{50}$) value of AR-C66096 was determined to be 34.9 ± 2.9 nM, and its E$_{max}$ 41.9 ± 3.6% compared with 2MeSADP as 100%. A final concentration of 10 μM forskolin was used in the experiment. DMSO was used as a solvent for the stock solution of forskolin, AZD1283 and ticagrelor. The stock solution of AR-C66096 was made with water.



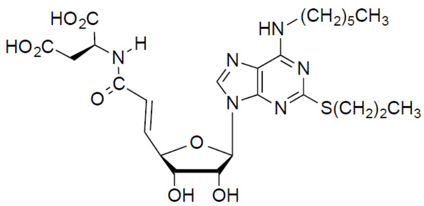
Extended Data Figure 7 | Docking models of different nucleotide analogues to the $P2Y_{12}R$ structure. **a**, The crystal structure of $P2Y_{12}R$–2MeSADP complex. **b**, Docking of 2MeSADP to the $P2Y_{12}R$ structure. **c**, Docking of ADP to the $P2Y_{12}R$ structure. **d**, The crystal structure of $P2Y_{12}R$–2MeSATP complex. **e**, Docking of 2MeSATP to the $P2Y_{12}R$ structure. **f**, Docking of ATP to the $P2Y_{12}R$ structure. **g**, Docking of AR-C66096 to the $P2Y_{12}R$ structure. **h**, Docking of AR-C67085 to the $P2Y_{12}R$ structure. **i**, Docking of AR-C69931MX (cangrelor) to the $P2Y_{12}R$ structure. 2MeSADP and 2MeSATP poses from corresponding crystal structures are shown in stick with orange and grey carbons, respectively. Docking was performed to the conformation of $P2Y_{12}R$ found in the 2MeSADP-bound structure, and the docked ligands are shown in sticks with purple carbons. AR-C66096 and AR-C67085 show the same interactions observed in the 2MeSADP complex. In addition, the C2-propylthio substituent of AR-C66096 and AR-C67085 is located in a hydrophobic pocket in proximity to helix IV surrounded by $F106^{3.34}$, $Y109^{3.37}$, $M152^{4.53}$ and $L155^{4.56}$. The γ-phosphonate group is directed towards helix III and interacts with $K80^{2.60}$ and $R93^{3.21}$. The C2 substituent and the γ-phosphonate group of AR-C69931MX show similar orientation as observed in the docking pose of AR-C66096 and AR-C67085. The N^6 substituent is directed towards helix VI in proximity to $Y109^{3.37}$, $Q195^{5.44}$, $F252^{6.51}$, $H253^{6.52}$ and $R256^{\ 6.55}$.



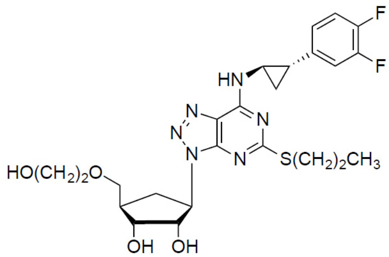
X = CF_2, AR-C66096
X = CCl_2, AR-C67085



AR-C69931MX (Cangrelor)



AR-C78511



AZD 6140 (Ticagrelor)

Extended Data Figure 8 | Ligands used in the docking studies. The chemical structures of parts of ligands that are discussed and used in the docking studies are shown. Ticagrelor and AR-C78511 could not be docked in a conformation similar to 2MeSADP because the presence of their bulky N^6 substituents would cause a steric clash with helices V and VI. AR-C78511 was previously shown to lack partial agonist properties.

Extended Data Table 1 | Data collection and refinement statistics

	P2Y12R-2MeSADP	P2Y12R-2MeSATP
Data Collection		
Number of Crystals usded	17	6
Space group	*C*222$_1$	*C*2
Cell dimensions		
a, b, c (Å)	65.1, 104.2, 169.4	75.7, 65.1, 100.7
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 95.5, 90.0
Number of reflections processed	335,625	26,125
Number of unique reflections	20,345	8,273
Resolution (Å)	50.0-2.50 (2.63-2.50) †	30.0-3.10 (3.27-3.10) †
R_{merge} (%)	19.4 (99.5)	22.2 (92.2)
$CC_{1/2}$	0.996 (0.587)	0.968 (0.413)
Mean $I/\sigma(I)$	12.4 (2.3)	5.6 (2.2)
Completeness (%)	100.0 (100.0)	92.2 (91.2)
Redundancy	16.5 (8.1)	3.2 (2.9)
Refinement		
Resolution (Å)	50.0-2.50	50.0-3.10
Number of reflections (test set)	20,345 (1,041)	8,273 (422)
R_{work} / R_{free} (%)	20.0 / 23.0	22.2/ 26.5
Number of atoms		
Protein	3,040	3,053
Ligand	29	33
Cholesterol	28	0
Lipids, PEG and waters	105	33
Overall *B* values (Å^2)		
P2Y12R	65.0	69.1
BRIL	59.3	146.4
Ligand	44.0	69.0
Cholesterol	87.8	n/a
Lipids and waters	70.2	84.9
RMSD		
Bond lengths (Å)	0.010	0.010
Bond angles (°)	1.03	0.87
Ramachandran plot statistics (%) ‡		
Favored regions	98.7	97.1
Allowed regions	1.3	2.9
Disallowed regions	0.0	0.0

The highest resolution shell is shown in parentheses.
† Values in parentheses are for highest-resolution shell.
‡ As defined in MolProbity.

Extended Data Table 2 | Binding affinities for different $P2Y_{12}R$ constructs

a

Constructs	[^{3}H]2MeSADP(K_d, nM)
WT	4.9±1.3
S83A	5.6±1.4
C97A	N.S.
R256A	16.1±6.2
C175A	N.S.
K280A	N.S.

b

Constructs	AZD1283 (K_i, nM)
WT	41.8±16.3
S83A	36.5±6.8
R256A	140±39

Affinity values of the agonist [^{3}H]2MeSADP determined in saturation binding to WT and mutant $P2Y_{12}Rs$ expressed transiently in COS7 cells (**a**) and inhibition by antagonist AZD1283 (**b**). Results are expressed as mean ± s.e.m. from 3–6 independent experiments performed in duplicate by methods described in ref. 5. N.S., not saturable or negligible specific binding within the radioligand concentrations used (0.4–46 nM).

CAREERS

COLUMN There could be good reasons for saying no to career moves **p.125**

@NATUREJOBS Follow us on Twitter for the latest news and features **go.nature.com/e492gf**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

SERGEY NIVENS/ECCO/SHUTTERSTOCK

PUBLIC RELATIONS

For your information

Scientific public-information officers need a gift for writing and a wide interest in all areas of science.

BY CHRIS WOOLSTON

Kurt Riesselmann remembers with delight the day that CERN researchers officially announced that they had finally detected what seemed to be the Higgs boson. He looked on as cheering scientists at Fermilab in Batavia, Illinois, watched a live stream of the announcement from CERN, the European particle-physics laboratory near Geneva, Switzerland. That moment, in the early morning of 4 July 2012, was especially meaningful for Riesselmann, who had earned his doctorate at the University of Wisconsin–Madison calculating theoretical Higgs interactions. But he wasn't just a highly interested spectator at the Fermilab viewing party. He had put the event together — and he had made sure to invite a *New York Times* photographer and a few key reporters. He wanted the US media to understand the significance of the particle, and he particularly wanted the world to know that Fermilab had played a large part in its discovery.

Riesselmann, head of the office of public information at Fermilab, is one of a growing number of scientists who have left research to become public-relations (PR) professionals, or, to use the term favoured by universities and national labs, public-information officers (PIOs). "We hesitate to use the word PR," he says. "We are not selling our science." As the Higgs news broke, he connected journalists with Fermilab scientists, wrote press releases and spread the word about the discovery and its significance in any way he could. "Even my neighbours were asking me about the Higgs particle," he says.

Government labs, universities and corporations around the world are eager to win publicity for their scientific endeavours, creating new job opportunities for scientists with a knack for communication. "It's definitely a growth industry," says Matt Shipman, a PIO at North Carolina State University in Raleigh. "Any university with a meaningful research programme is going to need people who can communicate science."

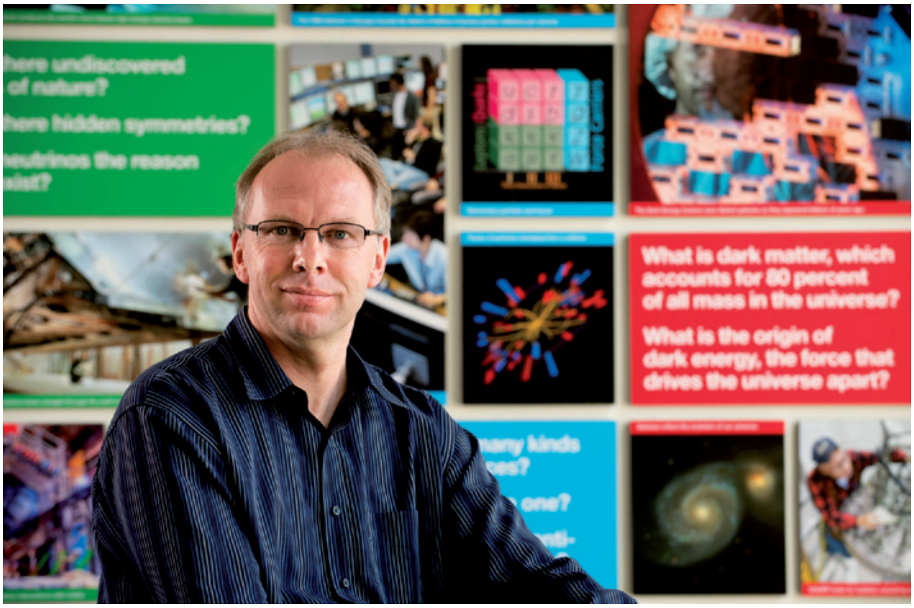
Opportunities for PIOs are global. In a 2013 survey, 642 members of the US National Association of Science Writers, based in Berkeley, California, said that they were staff writers for academic institutions, hospitals, private companies, government agencies or non-profit institutions, a nearly 20% increase from 2011. The European Geosciences Union, based in Munich, Germany, recently advertised for a communications officer with an advanced degree in either geosciences or journalism. The European Molecular Biology Laboratory in Heidelberg, Germany; a biotech company in Sydney, Australia; and King Abdullah University of Science and Technology in Saudi Arabia have all put out calls for PIOs or PR professionals with a solid scientific background.

Just about every form of communication can be part of the job. Videos, social media, blogs, press releases, conversations over the phone with scientists, investors and journalists — it's all in a day's work. "It's a rich, engaging environment," says Frank Orrico, a former molecular biochemist who is currently recruiting scientists to Element Scientific Communications, a Chicago, Illinois-based arm of the global PR firm Weber Shandwick. "I'm looking for people with a love of story-telling that goes beyond the academic stuff. If you want to get very deep into the science itself, this job may not be for you." His team currently includes 14 PhDs and two MDs based in the United States, London, Hong Kong and the Philippines. The list of clients is diverse, ranging from big pharma ▶

(including Roche and Eli Lilly) and biomedical-research groups (the Gladstone Institutes in San Francisco, California, and the multi-institution collaboration Orion Bionetworks) to the University of Iowa in Iowa City.

Those who make a successful career in scientific PR are often keener to talk about science than to do it. A PIO career was a good fit for Jonathan Wood, a media-relations manager at the University of Oxford, UK. He covers the university's vast medical-sciences division, which means he can be writing a blog post about the connections between smells and memories one day and a press release about cancer therapies the next.

Damage control can be less palatable; PR jobs often go beyond merely touting an institution's great work. Wood recently helped to craft the university's statement regarding an alleged case of scientific misconduct by a former DPhil student that led to the retraction of an article from the journal *Cell Metabolism*.



REIDAR HAHN/FERMILAB

Kurt Riesselmann left high-energy physics to handle public relations at Fermilab in Batavia, Illinois.

WRITING YOUR WAY IN

Wood has an undergraduate degree in physics from the University of Cambridge, UK, and a PhD in biology from the University of Leeds, UK, but his heart was never completely in the lab. "I realized that what I really enjoyed was journal club and presenting at meetings," he says. In 2006, he won FameLab, a prestigious communication competition in which scientists give live presentations in front of a panel of judges. "I get to work with some of the best researchers in the world, who really have a chance to improve health care," he says.

Not every scientist is cut out to be a PR professional, says science communicator and consultant Dennis Meredith, a biochemist by training and a former PIO at the California Institute of Technology in Pasadena, among other places. Meredith says that he asks potential PIOs a simple question: "Is writing a part of your basic personality? If not, this field isn't going to be very satisfying." Scientists also have to be able to think outside of their narrow topic of interest, he says, a talent that comes more easily to some than to others. "I personally always had intellectual attention-deficit disorder," he says. "You have to be prepared to have expertise that's a mile wide and an inch deep."

In the past, many scientists with a flair for writing and a promiscuous sense of curiosity ditched the lab for a job at a newspaper or magazine. But that particular escape hatch is narrowing fast as newspapers and magazines shed staff (see *Nature* **494,** 271–273; 2013). "The opportunity for a scientist to become a PIO is much better than for becoming a journalist, because that field has withered," Meredith says. Then again, he adds, a lot of seasoned journalists are looking to get out of the business, which means that scientists end up competing for PIO jobs against professionals with journalism degrees and stacks of bylines.

Meredith believes that scientists often deserve to have the upper hand in that match-up. Although many journalists go on to become highly successful and effective PIOs, they do have some limitations. Most obviously, Meredith says, scientists have much more experience of wading through academic papers to find the most interesting nuggets, even when they are working outside their particular area of research. Journalists may also have a dim understanding of the politics of a university research department, an area with which scientists generally are all too familiar.

"When a journalist becomes a PIO, it can be a little like a foodie becoming a chef," Meredith says. "They don't necessarily understand the internal institutional process." Unfortunately, he says, the university administrators who do the actual hiring do not always place much weight on scientific expertise. "A lot of vice-presidents say that they need to get a journalist in their PIO offices," he says. "It's up to the scientists to make the argument."

Meredith advises graduate students or postdocs who are interested in PIO work to offer their services to the public-affairs office at their university or their discipline's main organization. A few press releases or online stories could help a scientist to get a feel for the job while building a portfolio of writing samples. Orrico at Element Scientific says that he recently hired two PhDs largely on the strength of their science blogs.

Riesselmann got his start in communications in Germany, his native country, by doing

TRAINING THE COMMUNICATORS

Some courses on offer

- Imperial College London's one-year MSc in science communication prepares students for careers in public relations, journalism and similar careers.
go.nature.com/suspq8
- Bonn-Rhein-Sieg University of Applied Sciences offers an MSc in technology and innovation communications at its campus in Sankt Augustin, Germany.
go.nature.com/t1otva
- The two-year MSciComm course at the University of Otago in Dunedin, New Zealand, covers all forms of science journalism.
go.nature.com/xcddjl
- The master's degree in science and health communications at the University of Florida in Gainesville caters to scientists who want training in communication and journalists looking to specialize in science.
go.nature.com/dezv5d
- The Massachusetts Institute of Technology in Cambridge offers a one-year master's degree in science writing.
sciwrite.mit.edu
- The MA programme in science and health journalism at the University of Indiana in Bloomington includes a course in public-affairs reporting.
go.nature.com/artucl
- The one-year science-communication programme at the University of California, Santa Cruz, is open only to those with science degrees and at least some research experience. **C.W.**
scicom.ucsc.edu

occasional outreach work at a science museum during a postdoc at DESY, Germany's high-energy physics laboratory in Hamburg. "It's probably getting more and more difficult to get into the field the way I did," he says. Increasingly, he says, institutions looking to hire scientists for PIO positions will require candidates to have at least some formal writing training.

Many universities offer science-writing or communication programmes, giving scientists a chance to hone their writing skills and upgrade their résumés. Some programmes focus on journalism, but others offer specific instruction for PR and PIO work (see 'Training the communicators').

COVERING THE FIELD

Immediately after obtaining her doctorate in 2005 in materials science from the University of California, Santa Barbara, Aditi Risbud started the one-year science-communication course at the University of California, Santa Cruz. "I couldn't see myself loving research enough to make it my career," she says. "I knew I wanted to get into the PR field."

As part of her time at Santa Cruz, Risbud interned at the communications office of Stanford Medical Center in California, where she wrote press releases and stories about all sorts of medical research. After graduation, she was hired by Orrico at Weber Shandwick. Among other things, she was soon writing 15-page primers on tumour angiogenesis and treatments for multiple sclerosis. "I had to give myself an education in the medical field," she says. "I bought textbooks."

Risbud, who now works as a PIO in the College of Engineering at the University of Utah in Salt Lake City, recently wrote about electrical engineers who used an inexpensive inkjet printer to create microscopic 'plasmonic' structures that make it possible to use light beams to transmit data over metal surfaces. In an interview with Ajay Nahata, a professor of engineering and computer engineering at the university, Risbud got the type of quote that was guaranteed to get a reporter's attention. Plasmonic devices, Nahata told her, "have the potential to make wireless devices such as Bluetooth 1,000 times faster than they are today".

From medicine to engineering — that sort of versatility is the hallmark of a good PIO. "I was always intimidated interviewing scientists," says Meredith, "but then a very prominent astronomer told me that she was in awe of PIOs because we have so many areas of expertise. That was gratifying." ■

Chris Woolston *is a freelance writer based in Billings, Montana, and a graduate of the science-communication programme at the University of California, Santa Cruz.*

COLUMN

Choosing to stay

Saying no to a move is not easy, but there could be some very good reasons to stay put, says Yoshimi Rii.

As a graduate student contemplating a possible postdoc position, I have a constraint that I don't like to admit: I'd prefer not to move. Some may say that declining to be mobile is a cardinal career sin for young scientists. But I have my reasons.

ORLA/SHUTTERSTOCK

By the end of their doctoral programme, many students acquire not only research experience but also personal achievements: long-term partners, a set of great friends, perhaps a mortgage and maybe children. Yet many move — to another state, another nation, another continent — to pursue fellowships and research labs in the hope of positioning themselves for a future faculty job (see *Nature* **490,** 326–329, 2012). There's often more to their motivation than a sense of adventure — many universities shy away from hiring graduates from their own departments for new faculty positions. I have sat on faculty candidate-search committees and witnessed many qualified local candidates outcompeted by those who had ventured to labs far and wide.

If I were willing to be mobile, exotic and exciting opportunities would abound. At the Ocean Sciences Meeting in Hawaii in February, a researcher at the poster session invited me to do a postdoc in Chile. Although I entertained the idea for a few days, I knew that Chile wasn't going to happen. I have a husband who was recently made a supervisor in his job, and a great group of friends. But the real deal-breaker is that my mother and sister live here in Honolulu.

My mother is the healthiest 78-year-old I know, and yet, when I consider going away for even a year, I am paralysed by the fear of lost time. I hope to start a family in the next couple of years, and I want to maximize every moment my future family and I will have with my mum.

By choosing to prioritize geographical location for family reasons, I feel as if I am closing the door on academia. When I started to consider alternative career paths, I feared that my search for local career choices would severely narrow the scope, especially if I wanted to continue doing research. But I'm noticing that there are plenty of options if I think creatively about how to use my degree (see *Nature* **494,** 393; 2013).

With uncanny timing, I recently saw an announcement for a teaching chair at my old high school, a prestigious school with bright students. My former biology teacher is retiring after 52 years, and the school wants someone to implement a research programme in its new laboratory facility. I would be able to educate young students and instil in them a love for the ocean. I would get to stay 'at the bench' and stay at home — seems like a win–win. Still, I hesitated to apply.

First, my lack of exposure to alternative careers in graduate school makes them seem less worthwhile than a gloried professorship. I still hear people say, 'Oh, so-and-so won't make it in academia.' Such language perpetuates the unrealistic expectation that every PhD graduate should become a professor.

Second, part of me feels as if I'm making this decision for my family, and deliberately surrendering my future identity as an oceanographer or a scientist. I find myself fantasizing about an international postdoc or months at Palmer Station in Antarctica. Occasionally I wish I could become my adviser's mini-me.

In the end, choosing an alternative career is less about sacrifice and more about finally being in the position to take care of people other than myself. I'd like to give my mother some freedom to finally live her life, after she has heroically taken care of my special-needs sister for the past 40 years (not to mention a 34-year-old still in graduate school). I guess my feelings as a daughter supersede my drive as a young academic.

I decided to apply for the teaching chair. Nervous but excited, I hope for good news. Saying no to moving may close some doors, but it will open others, both personal and professional. ■

Yoshimi Rii *is a fifth-year PhD candidate in the School of Ocean and Earth Science and Technology, University of Hawaii at Mānoa.*

AN UNCANNY LIFE

A man out of time.

BY RYAN RUBAI

He has been here before. The stained peeling wallpaper, the dirt and debris marking the skirting boards are too unmistakable. Too real. He wonders if he should leave the room.

Opening the door, he plunges straight out into the abyss. It's pitch black but for a few puddles of bluish-black-like landing lights. A blinding headache engulfs him. He feels like he's being torn apart — ripped right down the middle, a real-life crease mark. Flashes of blue-tinged smoke erupt from the ether; crackling, hissing, screeching sounds come from nowhere and everywhere.

Bang!

He feels wetness trickling down the side of his head. The headache is gone. He's holding a gun. Gone. A subliminal from another time. He's in agony because he feels like this has happened already and it's like starting a film halfway through. He's on the floor but not on the floor. There's blood weeping from a head wound. There is no blood. There is no head wound. He's perfectly fine. Yet there he is, comatose on the floor with litres of blood pouring out. This is it, he thinks, the end of the line.

He is standing at the sink filling his coffee mug with water that froths over the sides. His foot bleeds. He cries out. He watches as his beard hair goes down the plug hole. The kettle boils. Shake twice, flush and wash. The cup slips out of his fingers.

It is evening now and he is sitting eating a microwaveable meal-for-one. He feels hungry and full at the same time and he can't figure out which one of him is the hungry one.

It was the first time he'd ever heard of it. Time Dissonance Syndrome. Reading from a medical dictionary, he found out that it was "a neurological disorder in which the patient perceives two or more similar but distinct lives simultaneously … Dr Hans Müller and Professor Henry Bohr first hypothesized that TDS shares similarities with the déjà vu virus (page 127) … see their respective entries for more information".

He knew there was something wrong with him when his father died. He was seven, in the backseat, his father hunched over the steering wheel and then time started to peel away, the inside of the car seemed to stretch. Then nothing. Just the sickening sound of metal on metal.

JACEY

The terrible thing was he had never been in a car crash and his father was still alive.

Years later, after countless sessions with different psychiatrists and doctors he was finally diagnosed with Time Dissonance Syndrome. The doctors told him that it was common to have more timelines as a child. As he grew older, they said, these parallel timelines would diminish. He found it difficult to explain his problem to people. So he stopped trying.

Every moment, even in sleep, he feels like he is living more than one life.

"You will experience some side effects."

"Like what?"

"Like … your perception of time will alter. You'll perceive time on a nonlinear level without the notion of causality. The symptoms only last, on average, three weeks — but …"

"But what?"

"That's based upon lineal time, whereas we have no concrete idea of when you'll perceive the effects to stop."

They're laughing at him. In the playground, pointing their fingers and laughing. Daring themselves to go nearer and nearer. They tease him. Torment him with their name-calling, whispering behind his back.

The school has sent a letter home. Inattentive, they say. In a world of his own. His parents try to understand, try to see it from his point of view but he doesn't seem to hear. They take it the wrong way. "Can't you see that there's something wrong with the boy? This is your fault, you should have been firmer with him."

"Tom? Tom, are you listening? If you don't want to learn, then you can just leave the classroom." He is not paying attention. He is disconnected. He has been in this scene before.

In his bedroom, he's curled up into a ball, squeezing himself ever tighter. His mouth clenched, eyes like slits. The pain in his head is intense. He can feel liquid trickling down the side of his head. It's dripping off his earlobe like the last gasps of the water tap. The pain stops. As quickly as it came. He feels his head. Nothing there. He lies on his back, feels around. His hand touches something. He knows what it is, doesn't want to look at it, can feel the shape of it. Cold and harsh. He can feel the weight of it, the contours between his fingers. The feel of the trigger. It's all there. He looks down at his hand. Nothing there, nothing at all.

He is ten.

He has been here before. He desperately wants to catch up to some future version of himself and hope he can feel normal.

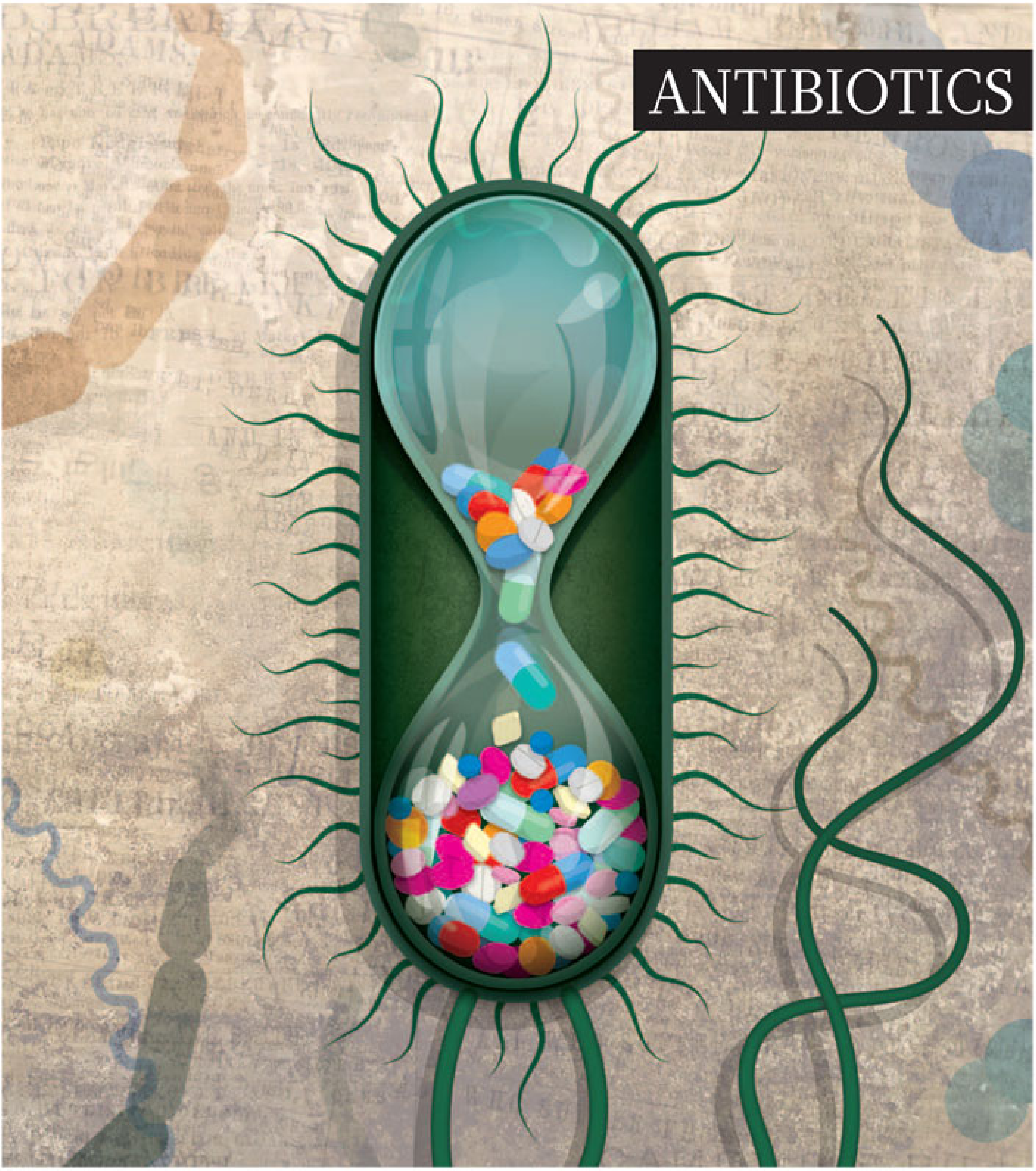
He has done it before. Felt the exploding headache that engulfs him. He has seen the smoke curl around the room.

He puts the gun up to his head.

He has been here before. ■

Ryan Rubai *hails from Northern Ireland. He writes like he walks: a bit wobbly but he gets there in the end. He's also bad at writing bios.*

nature OUTLOOK



The race against resistance

nature OUTLOOK

ANTIBIOTICS

1 May 2014 / Vol 509 / Issue No 7498

Cover art: Nik Spencer

In the United States in the 1960s, a visit to the doctor usually entailed getting a shot of penicillin. It seemed that anything from hangnails to headaches called for an antibiotic "just in case". At the same time, most cows and pigs on even small farms received antibiotics regularly whether they were needed or not. With such carefree prescribing, we cannot be surprised that we face big problems with antibiotic resistance today. It will take the combined efforts of academic researchers, pharmaceutical scientists and government officials, as well as the public, to fight the growing danger of untreatable bacterial diseases.

Promising approaches are being developed, however. For instance, industry and government are forming partnerships to develop new antibiotics (page S4), and researchers are seeking a better understanding of the mechanisms that drive bacterial resistance (S6). Sometimes — as with phage therapy (S9) — moving forwards means returning to yesterday's treatments. Other scientists are exploring entirely new sources of antibiotics, from building nano-size structures to trawling the oceans for useful organisms (S10). Natural products collected through such searches can be chemically modified to attack a specific target (S13).

Addressing the growing health-care challenge will require more than new drugs, however. Advanced diagnostics will be crucial, from assessing environmental sources of potential infections to reducing the waste of antibiotics (S14). And governments and social organizations around the world are working to reduce the overuse of antibiotics in agriculture and medicine (S16). All this means that future generations will have a very different experience of antibiotics.

We acknowledge the financial support of Roche in producing this Outlook. As always, *Nature* has full responsibility for all editorial content.

Mike May
Contributing Editor

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK
Cite as a supplement to *Nature*, for example, Nature **Vol XXX**, No. XXXX Suppl, Sxx–Sxx (2014). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE
The *Nature Outlook Antibiotics* supplement can be found at http://www.nature.com/nature/outlook/antibiotics

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES
For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea:Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

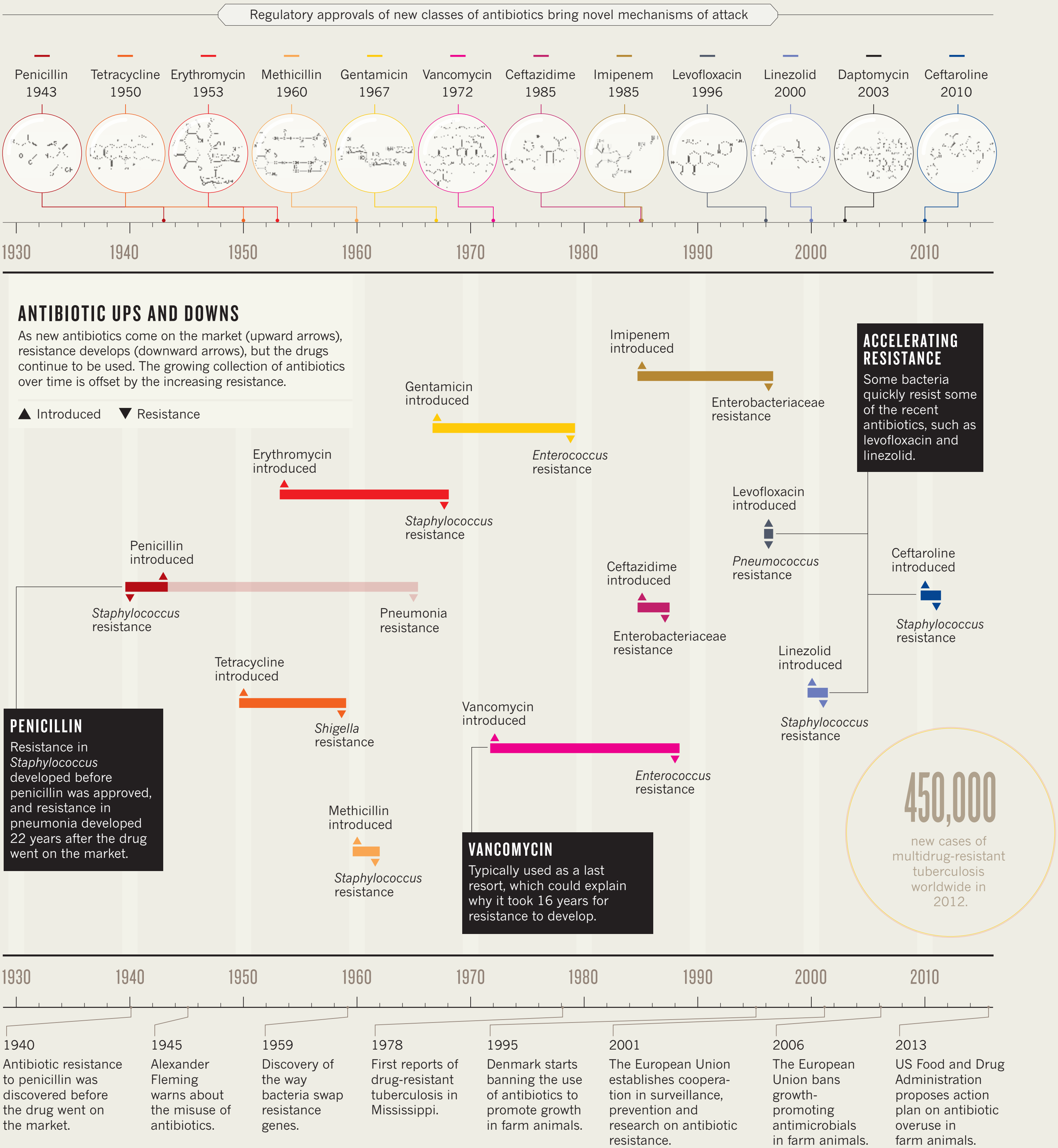
CUSTOMER SERVICES
Feedback@nature.com

CONTENTS

COLLECTION

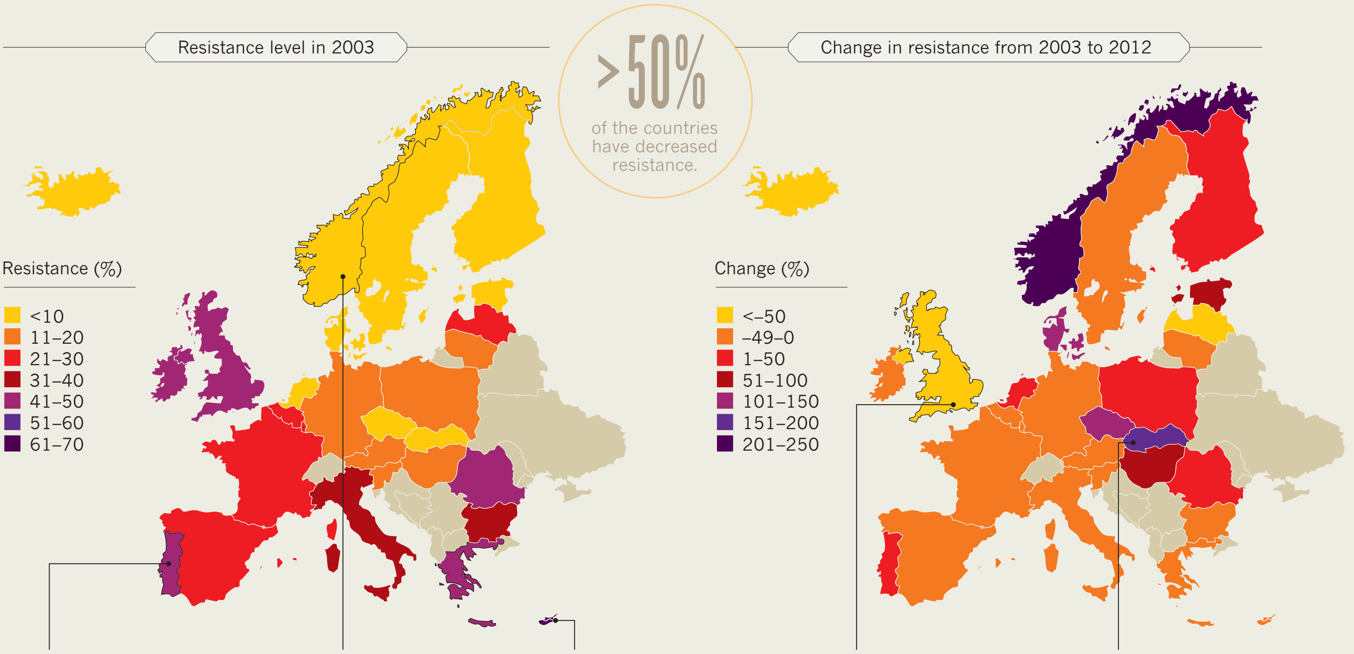
AN INFECTIOUS ARMS RACE

Winning the fight against infectious bacteria requires staying ahead of the organisms' uncanny ability to flank our frontal assaults. By **Karyn Hede**.



UPS AND DOWNS IN MRSA RESISTANCE

Over the past decade, many European countries have gained ground on methicillin-resistant *Staphylococcus aureus* (MRSA), but this deadly infection keeps spreading in others.



Portugal
With MRSA levels of nearly 50% in 2002, it spread infections beyond hospitals to public buses and other locations.

Norway
Despite MRSA levels of only 0.4% in 2002, it increased 225% by 2012.

Cyprus
Nearly 65% of infections were resistant to methicillin, the highest in Europe.

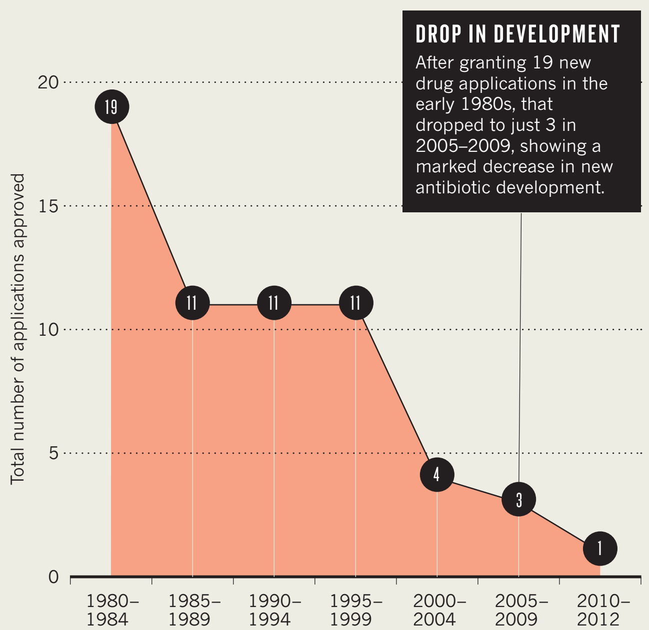
United Kingdom
Hospital policies that screen patients for MRSA could explain the decrease of nearly 70% from 2003 to 2013.

Slovakia
This increase of almost 165% supports a trend of increasing resistance in many countries that had low levels of MRSA infection in 2002.

SOURCE: European Centre for Disease Prevention and Control

A SHRINKING ARSENAL

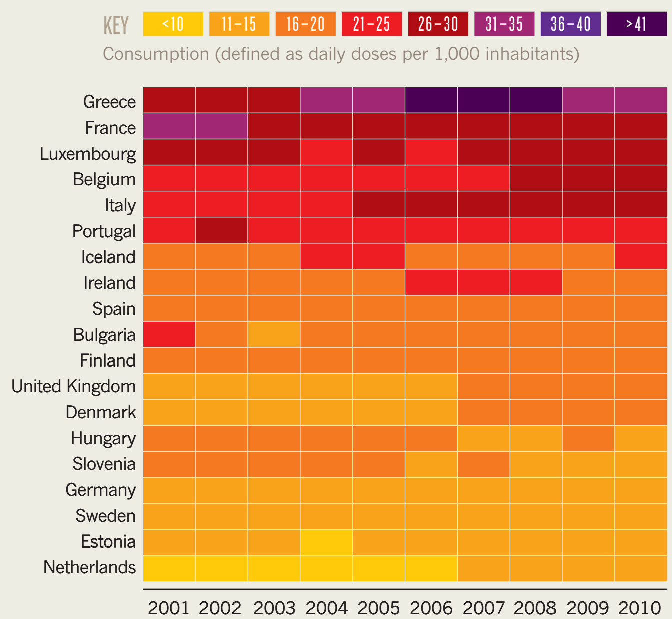
Since the early 1980s, the US Food and Drug Administration has approved fewer new antibiotics.



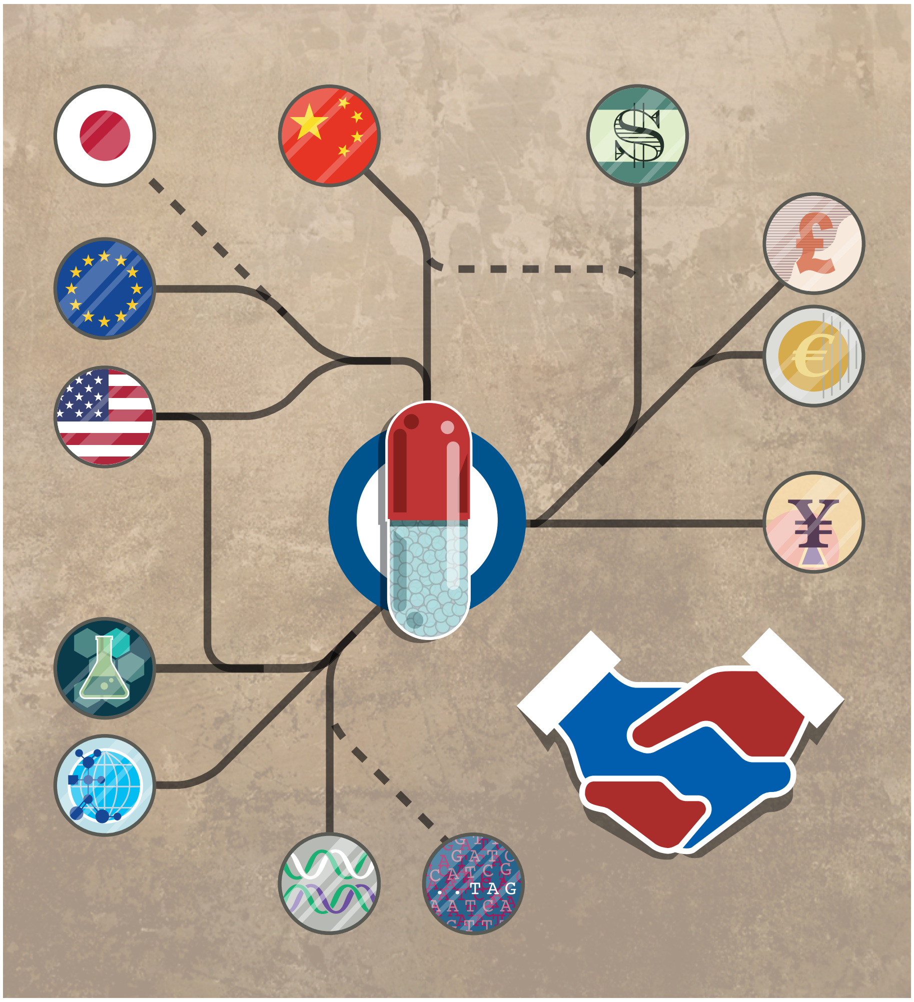
Source: US Centers for Disease Control and Prevention

USE OF ANTIBIOTICS VARIES DRAMATICALLY

Antibiotic prescribing practices in the European Union and European Economic Area vary widely, highlighting the need for standardized prescribing practices to reduce the spread of resistant bacteria.



SOURCE: European Centre for Disease Prevention and Control



NIK SPENCER

DRUG DEVELOPMENT

Time for teamwork

In the face of more drug-resistant bugs and fewer new drugs, partnerships promise a resurgence of antibiotics.

BY MIKE MAY

We are facing a bacterial-resistance catastrophe. In the United States, germs that are resistant to antibiotics infect at least 2 million people every year[1], making these infections more commonplace in the US population than cancer. At least 23,000 of those infected with antibiotic-resistant germs die. According to the US Centers for Disease Control and Prevention (CDC), intestinal infections with *Clostridium difficile* caused 400% more deaths in 2007 than in 2000, "in part because of a stronger bacteria strain that emerged"[1]. Likewise, the World Health Organization found that reported cases of multidrug-resistant tuberculosis in Africa soared[2] from 2,445 in 2005 to 18,146 in 2012 — an increase of nearly 650%. Daria Hazuda, who is head of discovery for infectious disease at Merck Research Laboratories in West Point, Pennsylvania, warns that "about half of today's infections are caused by strains of bacteria that are resistant to existing antibiotics."

Despite what Hazuda points out to be a "significant unmet medical need for new antibiotics", the pipeline for such drugs has slowed to a drip over the past 30 years. Between 1980 and 1984, the US Food and Drug Administration approved 20 new antibiotic drugs, but between 2005 and 2009 only three new antibiotics were approved. This crash in approvals reflects the plummeting productivity in the antibiotic-development sector of the pharmaceutical industry. Companies picked the easy fruit from the tree long ago, so new drugs have become harder to find. Also, the growing resistance in bacteria means that new drugs need to work in different ways, increasing the challenge even further.

The problem is rooted in economics. It takes hundreds of millions of dollars to develop any new drug. Recouping that investment is particularly hard to do with an antibiotic, largely because they are not valued highly enough to create a reasonable return. Brad Spellberg, who studies infectious diseases at the David Geffen School of Medicine at the University of California, Los Angeles, sums up the problem: "We will pay US$50,000 for a course of cancer chemotherapy that prolongs life by 3 months, but we don't want to pay more than $100 for a course of antibiotics that cures the target infection."[3] In the United States, insurance companies do most of the paying and often set the fees that can be reimbursed for specific drugs. Nonetheless, Spellberg suggests that the public are prepared to spend more on chemotherapy than on antibiotics. "People are terrified of cancer, but not of infections," he says.

The short duration of antibiotic prescriptions, usually lasting only a week or two at a time, makes it hard for these drugs to compete with the economic returns of medications that can be prescribed for life, such as treatments for diabetes. In fact, Merck's leading diabetes medication outsold its top antibiotic by about 7.5 times for 2009–2012 (see 'Diabetes far out-earns infections'). Such disparities could become even more pronounced as experts around the world preach the need to use even fewer antibiotics, prescribing them only when absolutely needed, to help slow down the development of resistance.

So, from a strictly rational business point of view, the pharmaceutical industry's retreat from antibiotics makes sense. David Payne, who leads the GlaxoSmithKline (GSK) antibiotics research unit in Collegeville, Pennsylvania, summarizes the issue: "I think the reason some companies move out of antibiotics is because there are other areas where they can get a greater return on investment."

STAY OR GO?

Pharmaceutical companies respond differently to the economic challenge of developing new antibiotics. For instance, Pfizer, which is based in New York City and was ranked by Forbes as the world's biggest drug company in 2013, closed its antibiotics research and development (R&D) facility in 2011. Apparently,

Pfizer has no plans to get back in the antibiotics business. Dean Mastrojohn, director of Pfizer's global media relations, says: "We will no longer conduct small-molecule R&D or business development activity aimed at in-licensing R&D-stage antibacterials and are seeking external partners for the few early stage programmes we had." So the world's biggest player seems to be set on staying out of antibiotics development.

Some of the pioneers in antibiotics plan to stay in this business, however. Merck, for example, was involved from the start, supplying the penicillin that was used for the first treatment of a patient with antibiotics in 1942.

Today, Merck hopes that advanced tools will speed up the development of medicines that fight infections. As Hazuda says: "We can use several technologies in high-throughput ways to enable us to figure out the mechanism by which new agents are working." Specifically, high-throughput sequencing, genomics and proteomics can reveal how an agent affects a specific bacterial target. "Then you can use that information to see if it's a new mechanism of action," Hazuda says. These tools might improve the profitability of creating new antibacterials. As an example, Merck has a monoclonal antibody, MK-3415A, in phase III trials for *C. difficile*, which the CDC blames for 14,000 deaths a year in the United States.

Roche closed its antibiotics division in 1999. "There was little reason to keep making new antibiotics then, because we thought we had more than we needed," explains Janet Hammond, head of Roche's Infectious Diseases Discovery and Translational Area in Basel, Switzerland. Recently, however, the company has started looking into this area again: "In the last year, we made a decision that we would go back into antibiotics," says Hammond. As part of its return, Roche plans to combine antibiotics research with other areas of expertise. "We are also strong in diagnostics," Hammond explains, and such tests could be used, for example, in clinical trials to more quickly assess the potential of a drug, potentially reducing the cost of development.

The increasing concern over infections has also spawned a range of antibiotics-focused companies. The list of specialists includes the following, among others: Achaogen in South San Francisco, California; Cempra in Chapel Hill, North Carolina; Hsiri Therapeutics in King of Prussia, Pennsylvania; and Qilu Antibiotics Pharmaceutical in Jinan City, China.

But even a specialist probably won't see a fast fortune from an antibiotic. More than a decade ago, Cubist Pharmaceuticals, based in Lexington, Massachusetts, received its first drug approval for Cubicin (daptomycin), which was the first of the so-called lipopeptide agents. It inserts itself into a bacteria's membrane and creates holes in it, killing the cell. This antibiotic can be used for serious infections, such as those caused by methicillin-resistant *Staphylococcus aureus* (MRSA). It turns out that making an antibiotic that treats such a dangerous infection, especially an infection that resists most existing medicines, can eventually generate big sales. In 2013, Cubicin gained blockbuster status by generating more than $1 billion in sales.

NATURE.COM
More about the economics of drug development:
go.nature.com/nTGLBQ

"The industry was dormant and we are using public–private partnerships to reinvigorate it."

Cubist used this commercial success to fund more research and acquire companies that could contribute to its antibiotics pipeline. In 2013, for example, Cubist acquired Optimer Pharmaceuticals and Trius Therapeutics. At the time of writing, Cubist has two antibiotics on the market and several more in clinical trials. As chief scientific officer Steve Gilman says: "We're leveraging our experience with Cubicin to develop new antibiotics." Such a concerted focus could be the key to success in this area.

EMPOWERED BY PARTNERS

Despite the economic success of Cubicin, the real key to energizing tomorrow's antibiotics arsenal could well come from public–private partnerships, in which government-funded departments or organizations provide financial and technical support to pharmaceutical companies. "We approached this looking at an industry that had become dormant, and we are using public–private partnerships to reinvigorate it," says Robin Robinson, director of the US Biomedical Advanced Research and Development Authority (BARDA). In particular, Robinson says that the US government provides a stable funding partner and technical expertise. "With antibiotics, these are major issues for industry."

As an example, Robinson's team works with GSK. "We are on a board with GSK to look at the entire antimicrobial pipeline and help them make decisions about how they allocate their money and how we allocate ours going forward," he explains. So rather than providing money and then walking away, BARDA has an ongoing advisory role in the research and regulatory decisions related to GSK's antibiotics pipeline.

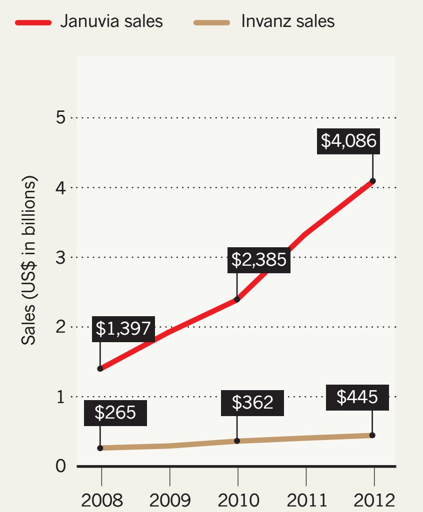
In Europe, the largest public–private partnership for developing new medicines is the Innovative Medicines Initiative (IMI), based in Brussels, Belgium. The IMI receives half of its funding from the European Union and half from pharmaceutical companies, so teamwork is baked into the cake.

The IMI recently launched a collaborative project as part of its 'New Drugs for Bad Bugs' programme, which is focused specifically on creating new business models for developing antibiotics. As Angela Wittelsberger, scientific project manager in charge of antibiotics at IMI, explains: "This brings stakeholders together who are not used to working together, including people from public health, academia, large and small industries, reimbursement agencies and government bodies." The pharmaceutical companies form the initial consortium, then the IMI invites proposals to select collaborators from the public sector, including universities, regulators and patient groups. This disparate collection of organizations will explore ways to efficiently and economically produce new antibiotics, hoping that teamwork will improve finances and innovation. Many international pharmaceutical companies — including AstraZeneca, GSK, Cubist and others — have already shown interest.

Although it is still in development, Wittelsberger sees a bright future for the programme. "The large realization from the pharmaceutical industry that this is something they should engage in and do collaboratively with their competitors all together is exciting," she says. Such a broad level of teamwork might be the best hope — maybe the only hope — of fighting the growing threat from antibiotic-resistant bacteria. ■

DIABETES FAR OUT-EARNS INFECTIONS

Merck's leading drug for diabetes (Januvia) outsold its leading antibiotic (Invanz) by more than US$11 billion over five years.



Mike May *is a freelance science writer and editor based in Lebanon, Ohio.*

1. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States, 2013* (2013).
2. World Health Organization. *Global Tuberculosis Report 2013* (2013).
3. Spellberg, B. *APUA Newsletter* **30**, 8–10 (2012).

MARK LOMOGLIO/ICON SMI/CORBIS

American football felt the fear of dwindling antibiotic effectiveness when methicillin-resistant *Staphylococcus aureus* infected the Tampa Bay Buccaneers.

MICROBIOLOGY

Resistance fighters

Science goes back to nature to decipher and disrupt the mechanisms by which germs evade antibiotics.

BY BILL CANNON

On Sunday 20 October 2013, American football team the Atlanta Falcons defeated the Tampa Bay Buccaneers. The game that day in the Georgia Dome, in the heart of downtown Atlanta, will not be remembered for its excellence — both teams finished at the bottom of their league. It was what happened after the game that hit the headlines. Another team —wearing not football uniforms but white hazardous-materials suits — entered and thoroughly disinfected the locker room used by the Buccaneers.

Back at their training facilities in Florida, several of the players had contracted methicillin-resistant *Staphylococcus aureus* (MRSA), a drug-resistant bacteria that is becoming increasingly prevalent — the number of people admitted to hospital in the United States with the bug doubled in just five years[1]. Falcons' officials were worried that MRSA had hitched a ride to Atlanta on the skin of their opponents.

The concern goes beyond the Buccaneers. MRSA and other *Staphylococcus* infections are so common that schools and colleges in the United States are routinely watching for them. The Chapel Hill High School in North Carolina, for instance, warns young athletes in its wrestling manual to "inform a coach immediately if they notice unusual marks on their skin", adding: "It is vital that the situation be dealt with quickly to prevent the infection of other team members." Almost everyone at the school knows someone who has been infected with MRSA.

However, our greatest risk of antibiotic-resistant infections may come not from fellow humans, but from cows, pigs and chickens. In 2011, about 80% of antibiotics in the United States went to livestock, to keep them healthy in intensive breeding conditions where infections are common and can spread quickly, and to boost growth (see 'The politics of antibiotics', page S16). Living within a mile of a factory pig farm in Iowa, America's top pork-producing state, nearly triples the odds of MRSA colonization[2]. In the United Kingdom, the expansion of pig farming in Staffordshire and Derbyshire has stirred fear of runaway antibiotic resistance.

In recent years, academic researchers have made leaps in understanding how microbes such as MRSA, carbapenem-resistant Enterobacteriaceae and others spread their resistance so far and wide, passing on mechanisms for evading antibiotics in a metaphorical genetic marketplace.

MOBILE RESISTANCE

Bacteria build resistance in two basic ways: vertically, by passing on mutations when a cell replicates; and horizontally, where genes are passed from one cell to another in DNA fragments called plasmids (see 'Two ways to spread'). "Plasmids are a dominant force in HGT [horizontal gene transfer]," says microbiologist Gautam Dantas of Washington University in St Louis, Missouri. "But there are important other ways in which HGT can occur, including transfer of linear parts of the chromosome from one cell to the other through the action of phages, as well as naked DNA uptake from dead cells," or DNA that is no longer encapsulated within a living bacterial cell and is free in an environment. Many researchers now consider horizontal transmission to be the primary driver of antibiotic resistance[3].

"Single genes give single phenotypes," says Dantas, referring to the observable characteristics and traits of an

NATURE.COM
Find out more about the development of antibiotic resistance:
go.nature.com/BrAFni

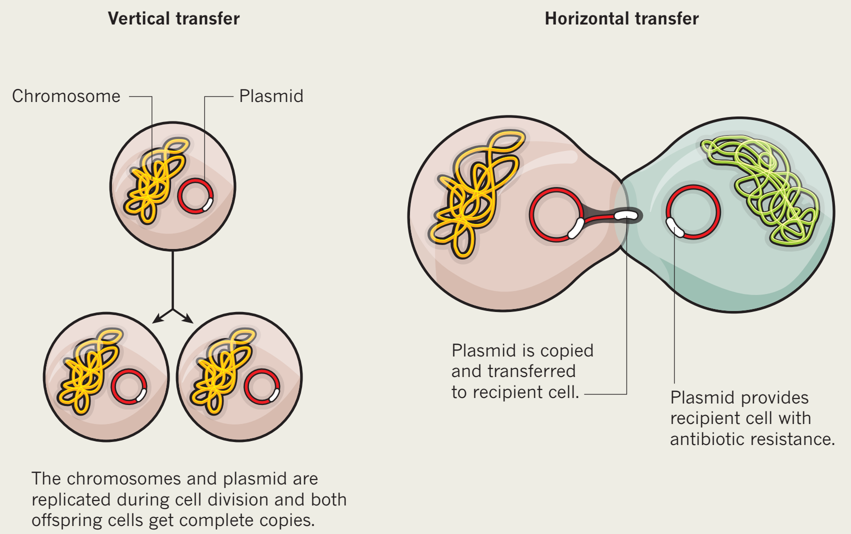
organism. "But we should not think that antibiotic-resistance genes are on their own." Any gene can be influenced by its neighbours, and genes often interact in clusters. Some will also be more mobile — better at jumping from one bacterium to another. Dantas suggests that metagenomics approaches — sequencing everything so that "we can identify resistance genes and what genetic context they are in" — will help to elucidate mechanisms of resistance, and distinguish those genes that are particularly mobile or that might work with other resistance genes to make a bug hard to target. As he explains, "This allows us to potentially distinguish resistance genes that are not going to be part of the clinical problem from that tiny fraction of genes that probably are."

Dantas views the concept of 'genetic epistasis', in which two or more genes influence one another, as strength in numbers. "We've seen, for instance from evolutionary studies, either retrospectively or prospectively, that particular genes that code for resistance tend to cluster together," he says.

By cataloguing resistance genes and the ways in which they interact, Dantas hopes that scientists can begin to make predictions. For example, how likely is a particular resistance gene to move from an organism that causes disease to team up with one that usually does not? "What's the likelihood these genes will give a particular phenotype in a particular organism?"

TWO WAYS TO SPREAD

Vertical and horizontal transfer mechanisms can spread resistance to antibiotics in bacteria.



SWITCHING IT UP

The topic of antibiotic overuse is not new; it is often implicated as the cause of resistance. So it seems logical that decreasing the use of antibiotics would increase antibiotic sensitivity. A few years ago at Uppsala University and Växjö Hospital in Sweden, evolutionary microbiologist Dan Andersson and collaborators tested this idea. Antibiotic use can either be reduced in general — that is, by cutting down on the total number of prescriptions or the length of treatments — or else the use of specific antibiotics can be controlled.

Andersson's team tried the second approach, letting a group of patients use one antibiotic until high levels of resistance developed, and then switching them to a different drug with the hope that resistance to the first would drop. The trial failed. "Generally, I think it's difficult to reduce resistance by reducing antibiotic consumption," Andersson says. "Over a long time, it's going to happen. But in many cases, it won't happen fast enough. For it to be useful to us from a medical standpoint, reversibility has to happen within a few years." Once resistance gets started, it is hard to stop.

The length of time it takes to reverse the problem differs between bacteria and antibiotic resistance, Andersson points out. Most mechanisms that cause antibiotic resistance have negative consequences for the germ, such as slowing its growth; this is known as a 'fitness cost'[4]. "The cost can be very small, but it's almost always there. There are some exceptions where actually the resistant bacteria are more fit than the susceptible ones even in the absence of antibiotics, and that's really worrisome." In these cases, Andersson says, "we wouldn't expect reversibility at all because then there is no cost to drive that." When resistance raises bacterial fitness, the problem is likely to spread, even without continued use of the antibiotic.

The combination of resistance plus an increase in fitness comes from a process called compensatory evolution. Andersson gives an example: "We have looked at a lot of antibiotics that affect protein synthesis. The antibiotics bind to the ribosome; if the resistance mechanism is due to a mutation in one of the proteins in the ribosome, the compensatory mutation might be in another protein in the ribosome."

> ***"We've seen that particular genes that code for resistance tend to cluster together."***

So the first mutation gives the bacteria its resistance, but also makes it a little less fit for survival; the second mutation makes the bacteria more fit without reducing its resistance to the antibiotic. In short, the second mutation compensates for the first in terms of the bacteria's ability to survive and reproduce. The ultimate result is bacteria that are antibiotic resistant and fitter than ever.

Overall, reversing resistance is not "something we should count on as a possibility for the future", Andersson concludes. "It's not going to save us. We need new antibiotics." He adds: "The industry and academic researchers have to get into action and develop new antibiotics. But that's easier said than done."

Equally difficult would be getting physicians to stop using drugs when the frequency of resistance in a given population reaches the level at which it ceases to ascend, and doctors would still need alternative drugs. Bruce Levin, a population and evolutionary biologist at Emory University in Atlanta, Georgia, and colleagues are working on a model that shows when the resistance levels off.

In the yet-to-be-published model, Levin and microbiologist Fernando Baquero of the Ramón y Cajal Institute in Madrid, Spain, and Pål Johnsen of the University of Tromsø in Norway, who studies the evolution of bacteria, suggest that if antibiotics are used until infecting bacteria reach a target frequency of resistance — say 20% of the bacteria acquire resistance to the drug — "we can keep the level of resistance to that drug at manageable levels," Levin says. "That would work well if we had an indefinite number of drugs." Of course, we do not. And not only is there a lack of new antibiotics, there is also the problem of multidrug-resistant strains, which can be resistant to drugs that have not been used for years. Stool samples collected by Levin's team at a daycare centre showed that about 25% of the *Escherichia coli* bacteria isolated were resistant to streptomycin, an antibiotic that, except for tuberculosis, hasn't been used as a treatment for nearly half a century. Yet the genes for streptomycin resistance are carried on a plasmid that is resistant to commonly used antibiotics. "In too many cases, resistance appears here to stay," says Levin.

COMMUNICATION BLOCK

New antibiotics might come from a better understanding of the structural elements and communications networks of bacteria. The

idea is to disrupt those components that lead to antibiotic resistance. A team led by Osamu Nureki, a structural biologist at the University of Tokyo in Japan, grabbed worldwide attention last year when it announced that it had solved the molecular structure of a multidrug and toxic extrusion (MATE) transporter, which confers antibiotic resistance[5]. In essence, the MATE transporter removes foreign things, such as toxins or antibiotics, from bacteria. While unveiling the crystal structure of the MATE transporter, Nureki's team also described a molecule that inhibits the MATE protein, the first hint of a way to block this mechanism of resistance and a possible target for drug developers (see 'Blocking an antibiotic rejection pump').

"There are lots of microbiologists trying to find bacterial Achilles' heels to make new classes of antibiotics," says molecular biologist Bonnie Bassler, a Howard Hughes Medical Institute investigator at Princeton University in New Jersey. Her lab, for example, is seeking ways to attack a bacterial communication network called quorum sensing, in which bacteria are able to determine when they are many, and in a strong position to attack. (John Woodland Hastings, a Harvard University photobiologist, first reported quorum sensing in 1970.)

In short, it works like this. Bacteria release molecules called autoinducers. If the concentration of bacteria is low, these molecules just float away. With more bacteria there are more autoinducers. "It's a concentration effect," Bassler says. "Every cell in the population is making and releasing its share of this signal molecule; the molecule increases outside the cell in proportion to cell number." She continues: "So when that molecule hits a threshold level concentration, the bacteria come into contact with those molecules and infer that they have neighbours around." The particular threshold, or 'quorum', depends on the species of bacteria.

What happens next, when pathogenic bacteria reach that quorum, can be ugly for the host. "The cells simultaneously turn on biofilm genes, virulence genes and toxins, and as a collective start releasing as a unit all of the compounds that make the host sick."

The fantasy, as Bassler bluntly describes it, is to block quorum sensing so that, even when they are crowded together, the bacteria cannot detect their neighbours and do not get the word out to build a biofilm — which is a tough-to-kill layer of bacterial cells — or to release toxins. "It looks promising," she says. "We can make molecules that are structurally similar to the real autoinducers, that jam the receptors of bacteria in a test tube, and they keep the bacteria from launching their virulence campaigns." As a result, she says, "We see that they can't make biofilms; we see that they don't kill tissue culture cells."[6]

> ***"The bacteria as a collective start releasing all the compounds that make the host sick."***

Nonetheless, this remains a dream, she says, because it is uncertain whether this would work when people typically seek medical attention, after developing an established infection — and after the bacteria have realized they have a quorum and have released their coordinated, virulent onslaught. Besides that, "we need a diversity of molecules that work at different points in the quorum-sequencing pathway, to learn which place is the best therapeutically," Bassler says.

Even if those molecules were developed, they would probably need to be improved and made more potent through medicinal chemistry. Only then could they be tested as treatments in animal models and eventually humans. It is likely that we are decades away from being able to exploit quorum sensing to develop new antibiotics.

DOWN TO EARTH

Like Andersson, Dantas and others, systems biologist Roy Kishony of Technion – Israel Institute of Technology in Haifa and Harvard Medical School in Cambridge, Massachusetts, envisages drug combinations based on a deep understanding of the interplay between drug interactions and competition between resistance and sensitive bacteria. Kishony and colleagues compare antibiotic resistance in natural environments and the clinic. Why, they asked, is the balance between antibiotic resistance and bacterial sensitivity — that is, susceptibility to antibiotics — maintained in nature while bacterial resistance quickly takes over in the clinic[7]? Kishony points out that nature developed antibiotics hundreds of millions of years ago to curb microbes that live in the soil. "We, as the human race, are fairly new in this business, and there's much we need to learn to uncover how antibiotic resistance works in nature," he says.

In nature, antibiotics "don't exist in isolation", Kishony says. "They exist in the context of many other compounds" and in an environment where there are multiple species of bacteria. "We've taken the antibiotic out of context," he explains. "We've removed the sets of checks and balances."

Resistance, Kishony explains, usually involves only a single or a few genes. "And many of these genes have been identified and worked out to a fairly good amount of detail mechanistically and molecularly," he says.

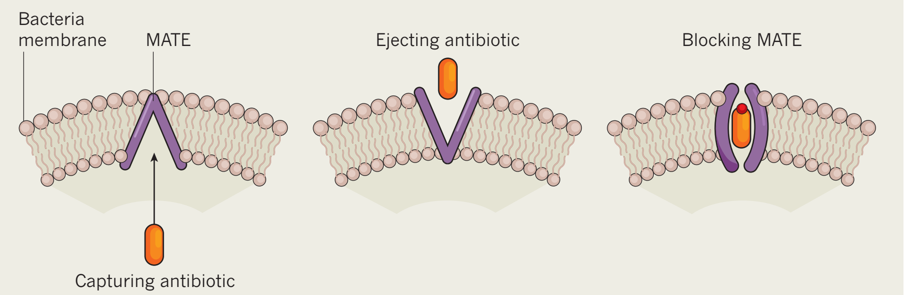
Kishony's team pays particular attention to the mechanisms in nature that counteract resistance, to identify candidates for clinical use. "We know that while a single drug always selects for resistance, some drug combinations can actually select against bacteria becoming resistant to any one of the individual drugs." As a proof of principle, Kishony cites a study out of his lab in which drug cocktails reduced resistance across a population of doxycycline-resistant *E. coli*[8].

He says there are likely to be many more similar mechanisms at play in the soil, which contains "an enormous number of different compounds that may interact in novel ways to inhibit the evolution of antibiotic resistance." Searching the soil for examples of these interactions may lead to new antimicrobial therapies that are more resilient to the evolution of resistance. As Kishony says, "We can invent things on our own, but we are quite likely to find great ideas and innovations just by looking in the dirt." ■

Bill Cannon *is a freelance writer based in Chapel Hill, North Carolina.*

BLOCKING AN ANTIBIOTIC REJECTION PUMP

Multidrug and toxic extrusion (MATE) protein removes antibiotics from bacteria, but it can be inhibited.



1. David, M. Z. *et al. Infect. Control Hosp. Epidemiol.* **33**, 782–789 (2012).
2. Carrel, M. *et al. Infect. Control Hosp. Epidemiol.* **35**, 190–193 (2014).
3. Dantas, G. & Sommer, M. O. A. *Am. Sci.* **102**, 42–51 (2014).
4. Andersson, D. I. & Hughes, D. *Nature Rev. Microbiol.* **8**, 260–271 (2010).
5. Tanaka, Y. *et al. Nature* **496**, 247–251 (2013).
6. O'Loughlin, C. T. *et al. Proc. Natl Acad. Sci. USA* **110**, 17981–17986 (2013).
7. Chait, R., Vetsigian, K. & Kishony, R. *Nature Chem. Biol.* **8**, 2–5 (2012).
8. Chait, R., Craney, A. & Kishony, R. *Nature* **446**, 668–671 (2007).

PERSPECTIVE

The age of the phage

It's time to use viruses that kill bacteria again, say **Shigenobu Matsuzaki**, **Jumpei Uchiyama**, **Iyo Takemura-Uchiyama** and **Masanori Daibata**.

Bacteriophages, or simply 'phages', are viruses that infect and in some cases destroy bacterial cells. Scientists started using phages as a medical therapy in the early 1900s. However, the technique fell out of favour in the 1940s, largely due to the introduction of antibiotics, which provided protection against a broader range of infections.

But as the problem of antibiotic resistance increases, more countries are revisiting phage therapy. Last year, for instance, the European Union (EU) funded a project called Phagoburn to explore the use of phage therapy to treat burn wounds infected with bacteria. Phagoburn involves institutions and hospitals in Belgium, France and Switzerland. The EU hopes the results from this project can be used, the project's website says, "for an optimization of current regulatory guidelines in phage therapy", because no new forms have been approved recently.

Countries that aim to introduce phage therapy will need to prepare their own guidelines for approving it, including methods for phage selection, preparation and administration. But recent advances in phage therapy suggest that such regulatory efforts would bring big rewards in treating bacterial infections.

AVOIDING IMMUNITY

Phages are a natural part of the microbial ecosystem. Environments such as sea water, fresh water and soil all contain millions of phage species. Different phage species are specific to particular bacterial species, and they can infect bacteria without harming animal or plant cells.

When faced with a bacterial infection, scientists first isolate candidate phages from the environment. The bacteria can be treated with a sample of water that naturally contains phages. If the bacteria die, the sample can be centrifuged, leaving the phages at the top to be collected and tested to see which ones killed the bacteria. Either the phage or its products, such as bacteriolytic enzymes called endolysins, can then be used as antibacterial agents in pills and ointments, often requiring just a single dose.

"ENVIRONMENTS SUCH AS SEA WATER, FRESH WATER AND SOIL ALL CONTAIN MILLIONS OF PHAGE SPECIES."

However, despite early success, phage therapy was largely abandoned when antibiotics came along, and is used today in only a few countries, including Russia, Georgia and Poland. Phage therapy declined in part because it focuses on treating specific infections, rather than on treating a range of bacteria. Some studies concluded that it failed because highly specific phages were simply tested against the wrong bacteria.

Yet phage therapy has several advantages over using antibiotics. First, because the bactericidal mechanism is completely different from the way antibiotics work, it is effective against multidrug-resistant bacterial infections. Second, phage therapy is highly species specific, meaning it is unlikely to change the bacterial flora of a patient and cause gastrointestinal side effects. Third, the propensity of phages to proliferate allows the use of very low doses.

Opponents of phage therapy often raise two potential problems: the appearance of phage-resistant mutant bacterial strains, and adverse reactions caused by the host's immune system against the phage. Modern techniques make it possible to address both of these concerns, however. First, using a cocktail of several different phages, or the advance preparation of mutant phages, overcomes any issue of bacterial resistance. Second, to stop phage therapy activating someone's immune system, a medical treatment can use phages with innate characteristics that are unlikely to elicit an immune response, use mutant phages that are not recognized by the immune system, or use some combination of the two. If a phage does somehow turn on the immune system, it can be treated with polyethylene glycol, for example, which will reduce the immune response.

Some phages can also produce toxins, but there is a way of resolving that problem too. Modern high-throughput techniques have moved phage therapy beyond screening water samples for potential treatments. Next-generation sequencing, for example, allows genomic DNA sequences from multiple phages to be analysed simultaneously. This makes it easier to detect suitable candidates for phage therapy that lack harmful genes, such as those that produce toxins or drug resistance.

THE SILK ROAD

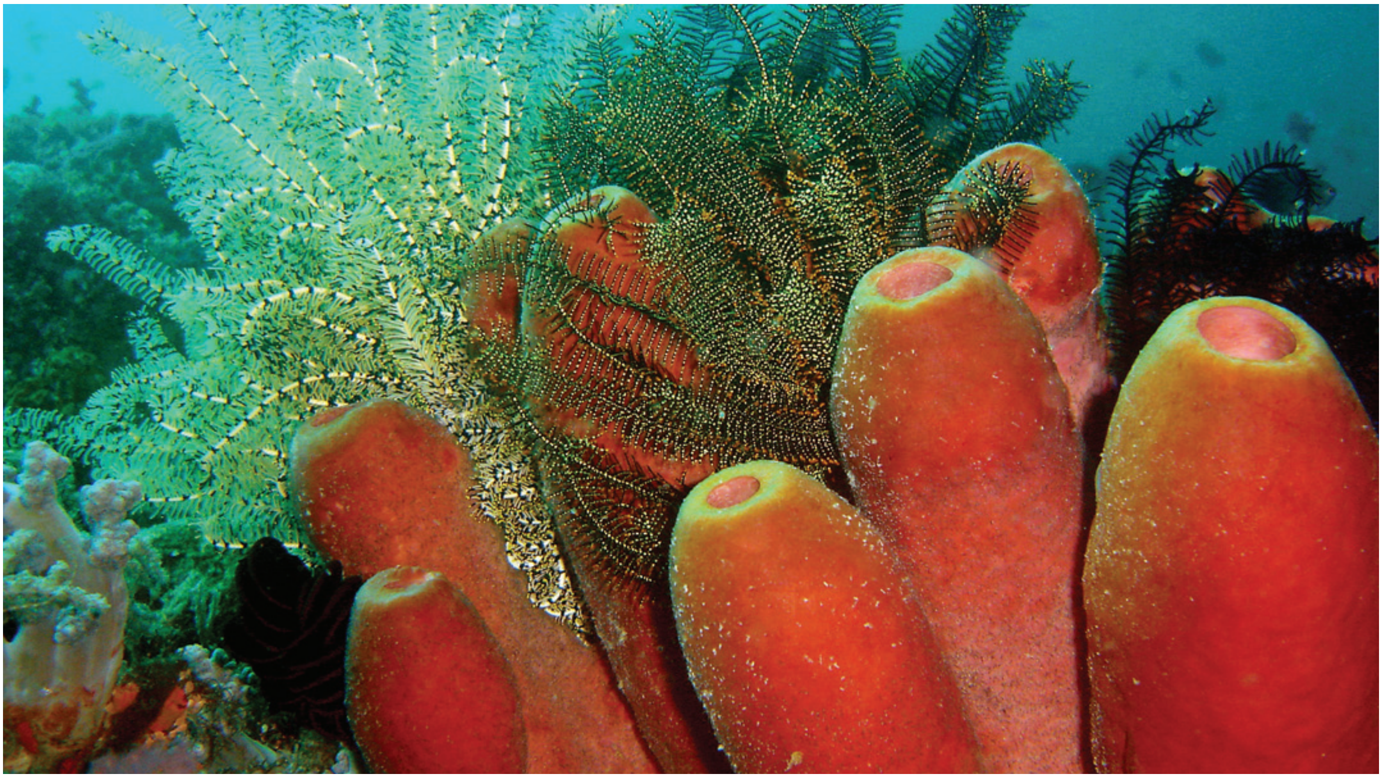
In addition, antibacterial research, like all drug discovery, benefits from lower-cost approaches. Modern drug development programmes often include studies on rodents, which require extensive experimental facilities and carry high experimental costs. For phage therapy, a less expensive invertebrate-based platform could be used. Kazuhisa Sekimizu and his colleagues at the University of Tokyo, for example, showed that silkworm larvae and mice provide comparable results regarding the effectiveness of experimental antibiotic therapies[1,2].

We have used silkworm larvae to test phage therapy against *Staphylococcus aureus* infections. Using two new phages to infect the bacterium, we found no adverse effects on the silkworm, but the phages did destroy the bacteria cells[3]. Our results using silkworm larvae were similar to those using these phages against *S. aureus* infections in mice.

Although we continue to obtain benefits from antibiotics, as we did in the twentieth century, the problems of antibiotic-resistant bacteria are set to increase, making it unlikely that antibiotics will remain effective forever. The latest techniques make it easier and faster to find phages to fight specific bacteria, and with less risk of resistance than using antibiotics. Phage therapy may be from a bygone era, but these advantages make today the age of the phage. ■

Shigenobu Matsuzaki *and* **Jumpei Uchiyama** *are associate and assistant professors of microbiology and infection, respectively;* **Iyo Takemura-Uchiyama** *is a clinical medical technologist; and* **Masanori Daibata** *is a professor of microbiology and infection. They are all at Kochi University Medical School in Japan.*

1. Hamamoto, H. *et al. Antimicrob. Agents Chemother.* **48**, 774–779 (2004).
2. Hamamoto, H. *et al. Yakugaku Zasshi* **132**, 79–84 (2012).
3. Takemura-Uchiyama, I. *et al. FEMS Microbiol. Lett.* **347**, 52–60. (2013).



Tomorrow's antibiotics might arise from unexpected sources, including this sponge, *Theonella swinhoei.*

DRUG DISCOVERY

Leaving no stone unturned

New antibiotic treatments could be found by combining novel and existing drugs, in drug-free nanoparticles, or at the bottom of the sea.

BY KATHARINE GAMMON

Under a microscope, the rod-shaped cells of the bacteria *Escherichia coli* look the same as ever. But there's a difference in the way they have been treated. Rather than giving them a growth medium containing the usual bonanza of nutrients, these bacteria face the nutrient-limited conditions they experience in nature. This simple change might provide a way to develop the next generation of antibiotics.

Unlike pharmaceutical scientists, who are governed by the odds of success, academic scientists have the freedom to carry out basic research in the hope it will lead to novel antibiotic agents. "What people struggle with in this area is how to do things in a fresh, new way," says Eric Brown, a biochemist at McMasters University in Hamilton, Ontario, in Canada. "There's not been a lot of success in recent years using modern drug discovery techniques." Brown realized that in order to beat *E. coli*, he had to think like *E. coli*, so he grows the bacteria in an environment that more closely mimics the inside of the human body during an infection — where paradoxically they are limited in vitamins and amino acids. It's work like this that has seen academic researchers rejuvenate the antibiotic pipeline.

Brown says he started to look for antibiotic targets in nutrient-limited conditions simply because no one else was looking there. First, he and his colleagues sifted through a library of 30,000 synthetic molecules, looking for any that can block the ability of *E. coli* to synthesize its own essential nutrients. About 300 molecules fit the bill. Then the researchers tested those molecules in nutrient-limited conditions, and 71 still worked. Ultimately, Brown and his colleagues identified three compounds[1] that might perform as antibiotics by blocking the bacteria's ability to make its own nutrients in an environment — like the human intestine — that lacks enough of them.

➲ NATURE.COM
For more on the search for new antibiotics:
go.nature.com/9DINjW

Such techniques might lead to fresh antibiotics — maybe many of them. Better still, the tool box includes a wide variety of strategies that could tackle the problem of antibiotics that are losing their power.

REVIVING FADED DRUGS

Instead of creating entirely new drugs, scientists might be able to reinvigorate antibiotics that have ceased to be effective by combining them with new agents, creating revamped compounds to refresh the drug pipeline. Robert Hancock, a microbiologist at the University of British Columbia in Vancouver, Canada, says that new approaches can breathe life into existing antibiotics. For instance, he created antimicrobial peptides called innate defence regulator peptides, which stimulate normal immune cells to sweep in and consume invading pathogens, and also suppress inflammation[2]. These peptides seem to be effective against a wide range of infectious bacteria, and are unlikely to lead to resistance because they have no direct effect on microbes.

Hancock says that although the peptides work as a stand-alone therapy, they are more likely to be used as an adjuvant to improve the

STEPHANIE SCHULLER/SCIENCE PHOTO LIBRARY

Starved *Escherichia coli* bacteria (green) challenged with synthetic molecules could lead to new antibiotics.

effectiveness of existing antibiotics. To explore the creation of treatments based on these peptides, Hancock started a company, now called Soligenix, based in Princeton, New Jersey. One of these agents is now in phase 2 clinical trials.

Another method for resurrecting fallen antibiotics is to alter their structure. Researchers at St Jude Children's Research Hospital in Memphis, Tennessee, changed the chemical structure of an existing antibiotic, spectinomycin, to create a class of antibiotics called spectinamides that are effective against new targets. The original antibiotic was used primarily to treat gonorrhoea infections, whereas the new class of drugs is effective against other infections including tuberculosis (TB) — an illness with a pernicious drug resistance problem (about 4% of new TB cases are resistant to multiple drugs, and about 20% are in people already treated for TB). In trials in mice with both active and chronic TB infections, one particular spectinamide — an analogue known as 1599 — was as good as or better than current TB drugs at reducing levels of the bacteria in the lungs. In addition, 1599 caused no serious side effects[3].

To create the spectinamides, the researchers designed complex three-dimensional models to examine how spectinomycin binds to the TB ribosome, where proteins are synthesized. By disrupting the ribosome, the drug was able to stop the bacteria from synthesizing proteins that it needs to survive. Because spectinamides bind to a different part of the ribosome to other drugs, they can be used in conjunction with existing therapies.

Richard Lee, a chemical biologist at St Jude's and the lead scientist on the 1599 project, says that using a chemistry-first approach led to some interesting developments. The new class of drugs can stop efflux, which some bacteria use as a defence against antibiotics, pumping the medication out of their cells — overcoming efflux has been a goal for some time. "Because they overcome efflux, they may be more appropriate for treating chronic infections where bacteria grow rapidly," says Lee.

KILLING THE PERSISTERS

Instead of breathing fresh life into older drugs, some researchers are looking for ways to trick bacteria into killing themselves. Kim Lewis at Northeastern University in Boston, Massachusetts, for example, focuses on a class of cells called persisters inside methicillin-resistant *Staphylococcus aureus* (MRSA). Persisters account for only about 1% of bacterial cells but are deadly — they are often dormant but wake up after an antibiotic has run its course and wreak havoc. After years of work, Lewis and his colleagues created a peptide — acyldepsipeptide (ADEP) — that activates the dormant persister cells and triggers them to self-destruct by degrading proteins and forcing the cells to digest themselves. Resistance and persisters go hand in hand, but the researchers found that when the peptide was combined with traditional antibiotics, resistance was kept to a minimum. Lewis and his colleagues tested ADEP in the laboratory and in a mouse model of chronic MRSA infection, and found that the mice were free from infection within 24 hours and showed no side effects[4]. The researchers are now working on clinical trials.

"Antibiotic tolerance can be as big a problem as antibiotic resistance."

Lewis says that antibiotic tolerance can be as big a problem as antibiotic resistance. For infections such as cystic fibrosis, for which there is no cure, and infections in people with hip implants or artificial heart valves, antibiotics suppress the infection but do not kill all the bacteria. "The infection keeps coming back because the dormant persister cells are not killed by antibiotics," says Lewis. "The cells resuscitate and cause a relapsing infection." The US Food and Drug Administration has traditionally asked drug companies to test compounds only against rapidly growing cells, which is one reason why there are not many treatments for long-term infections.

A similar technique might reveal a faster way to treat TB, which typically requires several courses of antibiotics over a period of 6–9 months. Lewis says that potential new drugs aimed at triggering persister cells to commit suicide could treat chronic TB infections in 24–48 hours.

SOAKING UP INFORMATION

Sometimes researchers go to the ends of the Earth to find potential antibiotics. In the case of Micheal Wilson, a microbiologist at the Swiss Federal Institute of Technology in Zurich, Switzerland, the search took him to the bottom of the ocean. The ocean floor could yield many as-yet undiscovered natural products. "We know how fruitful our relationship has been with terrestrial microbes, so we believe the oceans provide a vast underexplored resource for new natural products," he says.

Wilson studies microorganisms that live on sea sponges, which he calls "the guts of the ocean" because they are teeming with microbiota. Because sponges can't move, they have evolved a diverse range of adaptive relationships with bacteria. Wilson and his colleagues found that thousands of different bacteria live on the surface of one particularly productive sponge, *Theonella swinhoei*, but just a single species of bacteria from the genus *Entotheonella* produces nearly all the sponge's biologically active compounds[5].

The work is still at an early phase, but Wilson says that he and his colleagues are using a 'metagenomic' approach to identify genes, associate them with known compounds, and develop biochemical systems to produce potential drugs. In addition, other sponges could be studied to find bacteria that produce compounds that might be turned into medicines.

CLEANING UP TOXINS

Other researchers are thinking small. At the University of California, San Diego, nanoengineer Liangfang Zhang has been developing

nanosponges that fight bacterial infections. The technique focuses on attacking pore-forming toxins, which are proteins produced by bacteria such as *S. aureus*. These toxins bore holes in cell membranes, essentially causing leaks that disturb a cell's normal function.

"For bacterial infections, it's not always necessary to use antibiotics to kill the bacteria," says Zhang. Instead, a treatment could target the bacterial toxins, which are the source of all the negative consequences of an infection. If there is an effective way of removing the toxin, the bacteria are disarmed and the immune system can finish the job of killing the bacteria.

Each nanosponge is about 100 nanometres across and is made up of a biologically compatible polymer core wrapped in segments of natural red-blood-cell membrane. The nanosponges — disguised as red blood cells — create a decoy that collects the toxins so they can't harm the host's cells. The sponges have a half-life of 40 hours in tests on mice, and are eventually safely metabolized in the liver, along with the toxins. There is no antibiotic, so there is no opportunity for antibiotic-resistant cells to grow. Unlike other anti-toxin treatments, the nanosponges can work against a variety of different toxins, from MRSA to bee stings and snake venom — anything that creates pore-forming toxins.

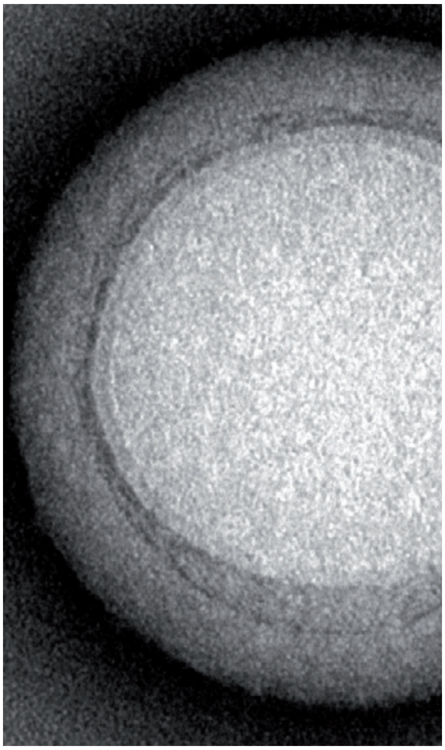
When the nanosponges were tested against alpha-haemolysin toxin from MRSA, pre-inoculation with nanosponges enabled 89% of mice to survive what would normally be lethal doses of toxin. Even administering nanosponges after the lethal dose led to 44% survival[6]. Further tests will combine the sponges with antibiotics with a view to making them even more effective.

Zhang is also working on a parallel nanosponge technique for vaccines that could prevent pore-forming toxins from taking hold[7]. He points out that there are more than 80 families of pore-forming toxin that this technology could disarm.

A SNIPER NOT A SHOTGUN

Instead of blasting someone with broad-spectrum antibiotics that can wipe out beneficial bacteria as well as disease-causing pathogens, it may be better to use a targeted approach. The biotech company AvidBiotics in South San Francisco, California, has created engineered versions of R-type bacteriocins, which are proteins used by *Pseudomonas aeruginosa* bacteria to kill other strains of bacteria by piercing their cell envelopes. AvidBiotics researchers say they chose this approach because the proteins are naturally species specific and easy to manipulate to new targets. The result is a protein-based drug called Avidocin.

So far, AvidBiotics has generated antibacterial proteins specific for a variety of bacterial pathogens, including *E. coli* O157:H7 (a strain that can cause fatal food poisoning), uropathogenic *E. coli* and *Clostridium difficile*. David Martin, AvidBiotics' chief executive, says the company has developed a large portfolio of proteins for



ZHANG RESEARCH LAB, UC SAN DIEGO

This nanosponge is made up of a nanoparticle wrapped in red-blood-cell membranes.

Gram-negative bacteria and a smaller one for Gram-positive bacteria. Clinical trials in humans are about a year-and-a-half away.

Martin points out that targeted attacks have many benefits. The protein against *C. difficile*, for example, is very narrow in focus and doesn't kill even related species. "The intent is that you can use it prophylactically, to prevent the disease rather than just treat the disease," says Martin. "These agents are so narrowly targeted that they don't disrupt the gut microbiota." They are designed to have a specific binding point and one dose is enough to kill the bacteria — a big advantage for avoiding potential side effects.

In hospitals and nursing homes, 3–10% of patients carry *C. difficile*, and Martin says they would benefit from treatment with Avidocin. The current use of broad-spectrum antibiotics can put them at risk of developing *C. difficile* colitis, which has a 20% mortality rate in people over 55 years of age. "Our strategy is to start with known bacterial pathogens, particularly in the gastrointestinal tract, and then eventually we want to be able to manage the gut microbiota using a sniper approach rather than a shotgun. Just kill the bad bugs," says Martin. "Don't kill beneficial bugs."

"No matter how successful the search for new antibiotics turns out to be, it might be best to use them in combination."

AvidBiotics is also collaborating with DuPont Nutrition and Health to create spray-on R-type bacteriocins that could fight pathogens such as *E. coli*, *Listeria* and *Salmonella* in packaged foods. Martin says his company is also creating narrowly targeted antibiotic alternatives to the drugs used in raising animals for food. In addition, the company has outlined a rapid-response platform with the potential to create targeted agents for use against emerging bacterial pathogens within days or weeks of acquiring the pathogen's genome sequence[8].

BETTER TOGETHER

No matter how successful the search for new antibiotics turns out to be, it might be best to use them in combination. Cancer, HIV and TB are all treated by combination approaches, and there's no reason why bacterial infections should be any different. Terry Roemer, who studies infectious diseases at Merck Research Laboratories in Pennsylvania, says scientists have long found that certain agents enhance each other's efficacy, for example using β-lactamase inhibitors to overcome bacterial resistance to β-lactam antibiotics. "The problem is, we're not really thinking rationally about which two antibiotics to pair that are synergistic together."

Roemer thinks that targets could be screened in a smarter, more rational way to create drug combinations that would decrease the likelihood of resistance. "For us, the mantra is: resistance is inevitable," says Roemer. "The best we can do is to try to reduce it as much as possible." Finding the best synergistic combinations of drugs would make it possible to do that.

As Roemer and colleagues have shown, MRSA infections were susceptible to β-lactam antibiotics when they treated the infection with two synergistic agents by containing resistance to a specific mutant that could be easily killed[9]. "So you can think of this as a situation where the potentiating agent restores the efficacy of the β-lactam, but resistance to that agent can drive two mutant forms of the pathogen that have a restored susceptibility," says Roemer.

In a world where effective antibiotics are becoming increasingly scarce, these approaches are set to benefit us all, from people receiving surgery to those with food-borne illness. And the advances are gaining speed. "Things have stagnated for a decade or more, but now it's a really exciting time," says Brown. "Pharma has been emboldened to take risks, but they still rely on crazy academics to do the wild things." ■

Katharine Gammon *is a freelance science writer based in Santa Monica, California.*

1. Zlitni, S., Ferruccio, L. F. & Brown, E. D. *Nature Chem. Biol.* **9**, 796–804 (2013).
2. Hancock, R. E. W., Nijnik, A. & Philpott, D. J. *Nature Rev. Microbiol.* **10**, 243–254 (2012).
3. Lee, R. E. *et al. Nature Med.* **20**, 152–158 (2014).
4. Conlon, B. P. *et al. Nature* **503**, 365–370 (2013).
5. Wilson, M. C. *et al. Nature* **506**, 58–62 (2014).
6. Hu, C.-M. J. *et al. Nature Nanotechnol.* **8**, 336–340 (2013).
7. Hu, C.-M. J. *et al. Nature Nanotechnol.* **8**, 933–938 (2013).
8. Scholl, D. *et al. PLoS ONE* **7**, e33637 (2012).
9. Tan, C. M. *et al. Sci. Transl. Med.* **4**, 126ra35 (2012).

PERSPECTIVE

MIKE LALICH

Synthetic biology revives antibiotics

Re-engineering natural products provides a new route to drug discovery, says **Gerard Wright**.

The drug discovery process often involves natural products or compounds that are inspired by them and modified by medicinal chemistry. For example, acetylsalicylic acid (aspirin) is one of the first cases of a natural product being modified to improve its drug-like properties. The ancient Greek physician Hippocrates prescribed concoctions made from the willow tree, which contains salicylic acid, to relieve headaches and other pains. Later, chemists modified and improved the properties of salicylic acid through synthetic reactions to produce aspirin. Today, synthetic biology is poised to reinvigorate interest in natural products as sources of new antibiotics.

By natural products, we mean genetically encoded small molecules that are the products of natural selection. As such, they represent privileged chemistry that has been tailored by evolution to interact with biological macromolecules, including proteins, nucleic acids, carbohydrates and membranes. It's not surprising then that more than 50% of drugs are either natural products or their derivatives. Many antibiotics, for example, have their origins in bacteria and fungi, which use natural products to interact with other microorganisms. Unlike most drugs produced in the lab, antibiotic natural products tend to be chemically complex molecules with intricate three-dimensional structures — properties essential for interacting with targets and for creating the antibiotic action. Sampling and expanding the chemical diversity of natural products is a critical component of discovering new antibiotics.

However, discovering antibiotics isn't that easy. For decades, the same molecules have shown up in screening assays designed to find compounds that kill bacteria. This suggests that perhaps millions of organisms must be screened to identify substantially different chemical scaffolds. In addition, antibiotics usually depend on elaborate chemical structures, complicating predictions of the likely drug qualities of a specific compound. These difficulties have greatly dampened enthusiasm for exploring new natural products.

> "TO APPLY SYNTHETIC BIOLOGY TO ANTIBIOTICS, ALL THAT WAS NEEDED WAS SOME CODE TO GET THESE FACTORIES TO WORK."

SYNTHETIC SOLUTION

Help is at hand, however. In the mid-1980s, David Hopwood, emeritus fellow at the John Innes Centre in Norwich, UK, demonstrated that scientists can manipulate the genes responsible for producing natural products to generate new compounds. Scientists at Kosan Biosciences in Hayward, California — a company that had Hopwood on its scientific advisory board — later expanded the strategy by inserting these genes in microbes to create compound-making factories. This approach creates libraries of derivatives of natural products by manipulating and mutating these biosynthetic genes. This revolutionary work set the stage for using synthetic biology to empower antibiotic discovery.

Synthetic biology is the application of rational engineering processes to biological systems. It offers solutions to the difficulties of using natural products in antibiotic discovery. In the past, researchers relied on random mutations to create antibiotics from natural products. Now, biological systems can be harnessed in a highly directed fashion to generate new products based on an understanding of the way such systems are assembled from a series of 'parts', such as a natural product's biosynthetic genes, to generate 'devices', such as new production pathways.

To apply synthetic biology to antibiotics, all that was needed was some code to get these factories to work. That came in 2002, from researchers at the Wellcome Trust Sanger Institute in Cambridge, UK, also including Hopwood. The scientists published the genome sequence of the antibiotic-producing bacteria *Streptomyces coelicolor*[1]. Among the bacteria's 8 million base pairs and nearly 8,000 genes were more than 20 new clusters of genes that produce natural products, some of which might work as antibiotics. This work presents opportunities to supercharge traditional approaches to developing antibiotics.

The sequence of *S. coelicolor* revealed the genetic programs responsible for the biosynthesis of the known natural products produced by this organism, plus a great many other clusters of genes that seem capable of producing other compounds with antibacterial potential. Similar results have since emerged from the genome sequences of dozens of other organisms, demonstrating great potential for the use of synthetic biology to produce natural products. No doubt some of these biosynthetic clusters produce antibiotics, and now they can be found by scanning genomes, instead of screening millions of microbes.

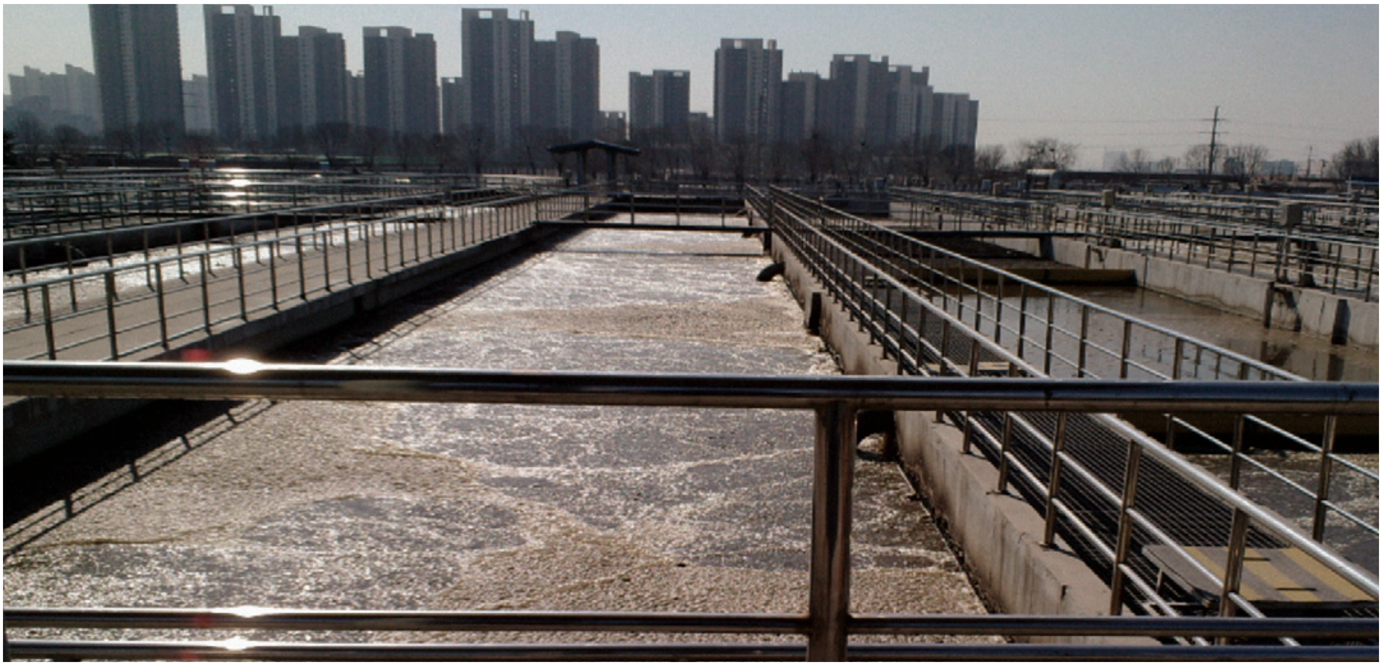
As an example, my colleagues and I used antibiotic resistance to enrich the collection of bacteria that produce glycopeptide antibiotics. This class of structures includes some well-known antibiotics, including vancomycin. By adding vancomycin to the process of isolating bacteria from soils, we can greatly enrich the collection of glycopeptide producers, as these must be resistant if they are to avoid suicide when they produce glycopeptides themselves. We developed high-throughput techniques that sorted out the structures that resembled antibacterials and eliminated 96% of the unwanted structures[2]. This process provides easy access not only to new antibiotics, but also to new 'parts' — biosynthetic enzymes — for synthetic biology.

The use of directed engineering of gene clusters that produce natural products offers a mechanism to greatly expand the chemical diversity of antibiotic chemical scaffolds. This enables the construction of chemical libraries suitable for screening and optimizing the promising compounds as drug candidates. Furthermore, transferring biosynthetic gene clusters into other host organisms might accelerate the production of desired compounds.

We are at the beginning of what might be a new era in antibiotic drug discovery that is enabled and fuelled by the application of synthetic biology to natural products. ■

Gerard Wright *is the director of the Michael G. DeGroote Institute for Infectious Disease Research at McMaster University in Hamilton, Ontario, Canada.*

1. Bentley, S. D. *et al. Nature* **417**, 141–147 (2002).
2. Thaker, M. N. *et al. Nature Biotechnol.* **31**, 922–927 (2013).



Enzymes that trigger antibiotic resistance can survive water treatment plants on China's Haihe River.

DIAGNOSTICS

Detection drives defence

Devices that quickly identify bacterial infections would benefit health and slow the spread of resistance.

BY REBECCA KANTHOR

In northern China, five large rivers and more than 300 smaller ones flow into the Haihe River, known as the 'mother river' of Tianjin, one of the country's largest cities. But something unpleasant has entered the river too: potentially deadly genes. In 2010, researchers from Nankai University in Tianjin reported finding genes in the Haihe River that can make bacteria resistant to antibiotics.

But it gets worse. A group of scientists from three universities recently found that New Delhi metallo-beta-lactamase (NDM-1) — an enzyme that can make bacteria resistant to a wide range of antibiotics — has survived the sanitizing processes of water treatment plants on the river[1]. The Haihe River exemplifies China's problems with antibiotic resistance and the need for improved diagnostics to track these life-threatening bacteria. Battling problems like this one, and the related health-care challenges, demands better tools for measuring antibiotic resistance around the world.

It may be no surprise to find antibiotic-resistant genes in China's waters, because China is the largest producer and consumer of antibiotics in the world. Wang Liping, a microbiologist at Shanghai's Jiao Tong University, describes the scene in most local hospitals: "You will see many people receiving an intravenous injection. If you have just a common cold but you have a higher temperature, you will be on antibiotics for sure."

Part of the problem is that antibiotics contribute to the financial stability of China's hospitals. Xiao Yonghong, who studies infectious diseases at Zhejiang University and directs China's Antibiotic Resistance Monitoring Network, says that antibiotics account for 20% of hospital drug sales, and that pharmaceutical companies offer incentives for physicians to prescribe them.

Better tests to help doctors diagnose infections could make a big difference to the way patients are treated in China, says Xiao. Faced with a patient with flu-like symptoms, many clinicians now prescribe broad-spectrum antibiotics as they wait several days for lab results to determine whether the infection is bacterial or viral. Advanced diagnostics, Xiao says, would help "to decrease the misuse of antibiotics". Diagnostic tools that speed up the identification of infections "would help doctors to make the right decisions about antibiotics".

Without these tools, resistance to antibiotics is increasing, and many so-called superbugs are on the rise, including methicillin-resistant *Staphylococcus aureus* (MRSA) and other difficult-to-treat microorganisms, such as extended-spectrum beta-lactamase-producing *Escherichia coli* and multidrug-resistant *Acinetobacter*. "During the past 10 years, antibiotic resistance rates have increased sharply in China," says Xiao. "This is very common in hospital settings." Resistance rates of *E. coli* and MRSA have more than doubled in China since 2000, he says, rising from 10% to 60%, and from 20% to 50%, respectively.

But China's problems with antibiotics are spreading beyond its borders. "With people travelling and tourism, resistance in one country could spread to other countries," Xiao says. "I don't think it is a local problem — it is a global problem."

A SIGN OF SEPSIS

As well as helping clinicians reduce the use of antibiotics, advanced diagnostics can also show when they are needed. In some cases, early diagnosis can make the difference between life and death. For example, sepsis is a whole-body infection, usually caused by bacteria, that is estimated to infect 4 million people a year, although

NATURE.COM
Recent work on rapid diagnosis:
go.nature.com/rJvTqg

global prevalence rates are uncertain. Sources indicate that at least 30% of people who get sepsis will die, often because treatment comes too late. According to the Sepsis Alliance in San Diego, California: "Sepsis should be treated as a medical emergency" requiring "rapid administration of antibiotics and fluids." The Alliance's website adds: "A 2006 study showed that the risk of death from sepsis increases by 7.6% with every hour that passes before treatment begins."

In China, sepsis in young children is a particular problem. Neonatal sepsis affects up to 15% of children up to three months of age in developing countries, and Chinese researchers have found[2] that "early diagnosis and treatment of neonatal sepsis is especially vital in China."

One way of diagnosing sepsis is to measure the level of procalcitonin (PCT) in the blood. In healthy people, high levels of calcium in the blood trigger the production of PCT, which is a precursor to calcitonin, a hormone that causes the body to reduce the calcium level. A bacterial infection, however, causes cells throughout the body to ramp up PCT production as part of the immune response. For a patient with sepsis, the level of PCT in the blood rises within a couple of hours of being infected and can be used to indicate the severity of infection. The difficulty is finding an easy way to measure it.

One rapid solution is the BRAHMS PCT-Q assay produced by Thermo Fisher Scientific in Waltham, Massachusetts. In this assay, a gold tracer binds to PCT so it can be quantified. One of the assay's benefits is its speed: in less than an hour, clinicians can distinguish between bacterial and viral infections, which do not raise PCT levels. During treatment, clinicians can also use the assay to determine when it is safe to end the course of antibiotics.

Improved diagnostics can also help clinicians assess the danger from previously ignored infections. For example, many countries track *Clostridium difficile*, a bacteria that infects the colon, but many Asian countries disregard it. A recent review of infections in Asia, however, concluded[3] that *C. difficile* infection "could be widespread in those regions where surveillance is currently lacking. Asia may be facing a 'perfect storm' as heavy usage of antibiotics combines with an ageing, increasingly hospitalized population."

One approach for detecting *C. difficile* is an assay developed by Quidel, based in San Diego, California. "The gold-standard culture method takes four days and a lot of work," says Timothy Stenzel, Quidel's chief scientific officer. "What patient wants to wait four days to determine the cause of their potentially life-threatening diarrhoea?" Instead, he says, the Quidel test provides an answer in just two hours. Slowing down a bacterial epidemic takes that kind of speed, but there are many more bacterial infections than sepsis and *C. difficile*.

> ***"Instead of identifying just one infection, tomorrow's tools could provide broader diagnostic possibilities."***

A BROADER BASIS

Instead of identifying just one infection, tomorrow's tools could provide broader diagnostic possibilities. Ralph Weissleder, who studies inflammatory diseases and is director of the Center for Systems Biology at Massachusetts General Hospital, uses genetic markers — parts of the bacterial genome — to detect bacterial infections. These markers can identify virtually any bacteria. This is particularly difficult when bacterial numbers are low. "As with most of the bacterial detection devices, the challenge is always detecting a few 'bad' bacteria in a sea of 'good' bacteria," Weissleder says.

So far, the team has tested the technology by identifying tuberculosis[4], searching for gene sequences unique to *Mycobacterium tuberculosis*. A patient's blood sample, or other body fluid, is placed into the device, which uses the polymerase chain reaction to make more copies of bacterial DNA, making it easier to identify (see 'Infection-detecting device'). Weissleder hopes the kit will be accessible to doctors all over the world, including in underdeveloped regions. The kit is little thicker than a microscope slide and the sensors are inexpensive to make. "In two hours you know exactly what kind of bacteria are in there and what the resistance profiles are." This technology can also identify many other bacteria, provided they have unique gene sequences. T2 Biosystems, a company co-founded by Weissleder, is developing the technology into a medical device, which is in phase III clinical trials.

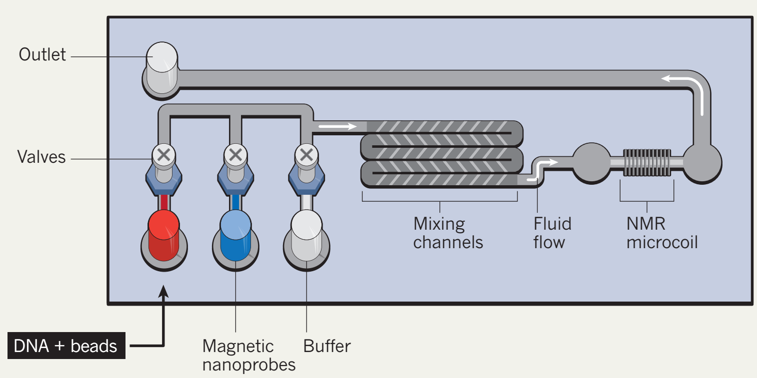
Success in one medical arena can come from failures in another. For 25 years, Nicholas Topley at Cardiff University, UK, has worked with kidney-failure patients on a dialysis process using peritoneal fluid, which provides lubrication between the organs and body wall in the abdomen. Working with Matthias Eberl, an immunologist in his lab, Topley realized that peritoneal fluid contains evidence of infections that afflict dialysis patients. This means the pair could "look at whether the biomarkers we measure actually predict the type of pathogen," Eberl says. By seeing which biological markers in the peritoneal fluid have changed, Eberl and Topley hope to find a 'fingerprint' that identifies each kind of pathogen. They have already found that analysing the peritoneal fluid can distinguish some types of bacteria[5].

As well as helping patients with kidney failure, this technology could be used for bacterial infections in general. The team is collaborating with Mologic, a biotech company in Bedford, UK, to develop an at-home dipstick test for kidney-failure patients, which would appear and function similar to a pregnancy test. "It would gain precious time as people could basically diagnose themselves and see if they have an infection," Eberl says. "They could then go directly to the hospital and get the most appropriate treatment." This could improve the effectiveness of the treatment and reduce the chance of the bacteria developing resistance.

Improving the accuracy and speed of bacterial diagnostics around the world could have an enormous impact. The potential to save lives with faster and more targeted diagnoses, decrease unnecessary and often incorrect prescriptions, and even help identify early on where bacterial resistance could occur, will have a drastic effect on the way patients are treated. For these tools to do their jobs, though, they must be accessible to the doctors and patients who need them. As Liping says of using advanced bacterial diagnostics: "It can be a life-saving situation in many cases." ■

INFECTION-DETECTING DEVICE

In this microfluidic chip, bacterial DNA attached to beads is labelled with magnetic nanoprobes, mixed with a buffer in channels, and a nuclear magnetic resonance (NMR) microcoil identifies the infection.



Rebecca Kanthor *is a freelance journalist based in Shanghai, China.*

1. Luo, Y. *et al. Environ. Sci. Technol. Lett.* **1**, 26–30 (2014).
2. Li, Z. *et al. Int. J. Clin. Exp. Med.* **6**, 693–699 (2013).
3. Collins, D. A., Hawkey, P. M. & Riley, T. V. *Antimicrob. Resist. Infect. Control* **2**, 21 (2013).
4. Liong, M. *et al. Nature Commun.* **4**, 1752 (2013).
5. Lin, C.-Y. *et al. J. Am. Soc. Nephrol.* **24**, 2002–2009 (2013).

Agricultural livestock, such as these factory-farmed pigs, consume most of the antibiotics around the world, reducing the effectiveness of the drugs.

PUBLIC HEALTH

The politics of antibiotics

Policy-makers and medical experts need to think globally if we are to prevent an antibiotic 'tragedy of the commons'.

BY MEGAN CULLY

In the early nineteenth century, demand for wool in England skyrocketed, and shepherds expanded their flocks accordingly. At that time, many sheep grazed on communal fields known as the commons. But the large flocks overgrazed the commons, so there was not enough food for other sheep or for cows, which became small and stunted. Inevitably, all the users of the commons suffered because there simply wasn't enough grass to go around. This 'tragedy of the commons' is a classic example of how individuals — acting rationally and according to self-interest — can damage or destroy communal property. The idea is often used to frame environmental issues, but it also has resonance for bacterial resistance. Individuals who overuse antibiotics are like the shepherds with large flocks, and the resulting tragedy is the rise of resistance to antibiotics.

Repeated exposure of microorganisms to antimicrobial agents — in humans, animals or the environment — increases the selection pressure for antibiotic resistance. In many developed nations, livestock consume an estimated 50–80% of antibiotics. Humans consume most of the rest, with crops, pets and aquaculture collectively accounting for about 5%.

To curtail the antibiotics tragedy, a wide range of people and governments must make changes. Antibiotics must be used wisely and sparingly, and farmers, physicians and the general public must all recognize the importance of this valuable resource.

PUTTING MEAT ON THE BONES

Antibiotics have many uses in livestock. They are used to treat active infections, prophylactically to prevent infections, or even as growth promoters. It was shown in the 1940s that treating pigs and chickens with broad-spectrum antibiotics increased their size and the amount of meat produced. Antibiotics then became widely used as growth promoters, despite evidence as early as 1951 of antibiotic resistance in farm animals. In countries that still allow this type of use, including the United States, more than half of the antibiotics given to animals are for the purpose of increasing growth.

The use of antibiotics as growth promoters is most common in pigs. China produces almost 30% of the world's pork, and its farmers use four times more antibiotics than their US counterparts to produce the same amount of meat. Yong-Guan Zhu from the Institute of Urban Environment in Xiamen, China, has found that antibiotic resistance genes are enriched by up to 28,000-fold in manure and soil on Chinese pig farms[1]. These antibiotic resistance genes are easily transferred from bacteria that infect animals to those that infect humans. Zhu fears that overusing antibiotics in livestock could make a large contribution to antibiotic resistance. "We are increasingly developing good sewage treatment plants [for human waste] and they help mitigate the problem," says Zhu. "But the animal industry is less developed in coping with manure and waste water" (see 'Detection drives defence', page S14). Indeed, Stuart B. Levy, president of the Alliance for the Prudent Use of Antibiotics, a US organization that works to decrease antibiotic misuse, says that "some of the more unusual [mechanisms of resistance] have emerged among the Chinese."

Changes have already been made in Denmark. In 1995, the Danish minister for agriculture and fishing banned avoparcin — one of the 11 antibiotics used for growth promotion — because it was seen to increase antibiotic resistance. Where the Danes led, others followed, and the use of antibiotics for growth promotion was banned entirely in the European Union (EU) in 2000. The use of antibiotics in Denmark has fallen by half since the early 1990s, despite an increase of more than 60% in the number of pigs. The use of antibiotics in animals still varies widely among EU members, however, with Nordic and Baltic states using about 10% of that used by Spain and Italy for a given amount of meat[2] (see 'EU countries cutting back on antibiotics in farm animals').

"Originally, everybody in Denmark expected that this would have a negative influence on Danish pig production," explains Frank Aarestrup, a microbiologist at the Technical University of Denmark in Copenhagen who was instrumental in reducing the country's use of antibiotics. But the high productivity achieved in the early 1990s continued when antibiotics were banned[3] — and resistance to many of the 11 banned antibiotics more than halved in bacteria obtained from animals and humans.

The United States, where antibiotics are still widely used as growth promoters, uses about 1.5 times as many antibiotics as Spain, one of the heaviest European users. In December 2013, the US Food and Drug Administration encouraged the agricultural industry to stop using antibiotics as growth promoters, but Levy doesn't expect such a light touch to have much effect. He thinks that outlawing these drugs for growth promotion is inevitable, but he accepts it will take time. In Europe, it was possible to simply outlaw the use of antibiotics as growth promoters. "In the United States, you're going to have a recommendation that will probably have to go through a judge. It's a whole different process — much lengthier and more expensive and time-consuming," he explains.

THE RIGHT PRESCRIPTION

Antibiotics use in humans also varies by region. "The average Greek or Frenchman consumes about three times as much antibiotics as the average Dutchman or Swede," says David Livermore, a microbiologist at the University of East Anglia in Norwich, UK, who leads on antibiotic resistance at Public Health England (PHE), an executive agency of the UK Department of Health. The Department of Health's five-year strategy outlined seven key areas to address antibiotic resistance; PHE is charged with improving surveillance, optimizing prescribing practices, and educating doctors and the public about the risks of antibiotic misuse.

Anthony Kessel, director of strategy at PHE, was particularly concerned by a 2011 survey of prescribing practices at English hospitals, in which meropenem — a drug that can kill a broad range of bacteria — was the ninth most commonly prescribed antibiotic. Such broad-spectrum antibiotics are exceptionally useful because they can treat so many different types of infection. But their overuse could leave physicians with no useful drugs to target bacteria with multidrug resistance. "These drugs should really be held for use as a last resort, when everything else has failed," says Kessel.

Many patients, particularly children with sore throats in the United States, are given antibiotics they don't need. Viruses cause about 90% of sore throats — and antibiotics don't kill viruses. Levy thinks it is possible to cut US community-based antibiotic use by up to 50% by identifying which sore throats are caused by bacteria. The key to this is better diagnostics.

"We must also get better at optimizing the duration of antibiotic treatment," says Livermore. Courses of antibiotics could be tailored to individual patients, particularly in hospitals. Livermore thinks that some courses of antibiotics, which often last 7–14 days in the United States, could be shortened by 2–3 days if biological markers were used to determine whether an infection is still present.

For physicians, handing out prescriptions is easier than explaining antibiotic resistance during a five-minute consultation. But educating doctors and parents with sick kids on the risks of overusing antibiotics might just spark the enlightened self-interest shown by Danish farmers in the 1990s, who voluntarily gave up using all 11 of their antibiotics long before being mandated to do so.

WORKING TOGETHER

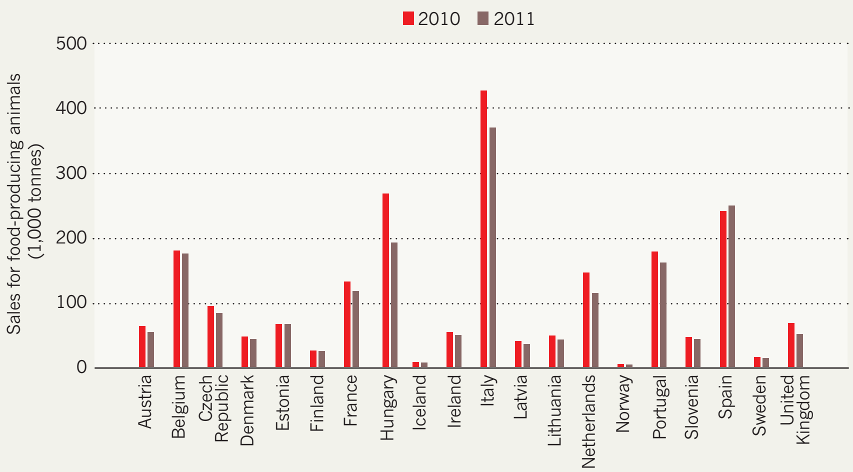
Education is also a central theme of Antibiotic Action, a UK-based organization headed by Laura Piddock. She and her colleagues have worked with UK and EU parliamentarians and the UK Department of Health, and the organization's 'antibiotic action champions' around the world are educating people about the importance of tackling antibiotic resistance. One of their key target groups is the public. Exposure to chemicals such as triclosan, the active ingredient in many antimicrobial cleaning agents, can select for mutations in genes that encode bacterial efflux pumps, which can eject antibiotics. Antibiotic Action targets women's magazines and lifestyle TV programmes to educate users of antimicrobial products about their potential harm. "Soap and water is enough," says Piddock.

There are many diverse approaches to reducing the use of antibiotics. For example, Aidan Hollis, an economist at the University of Calgary in Canada, suggests imposing a fee for the non-human use of antibiotics[4]. He says this would deter the low-value use of antibiotics, would be easy to administer, and would be easy to replicate in other countries. The fees would also generate revenue that could help to further our understanding of antibiotic resistance, discover new antibiotics, and educate more stakeholders.

Farmers, physicians and patients need to recognize the value of antibiotics and protect this vulnerable resource. In the absence of enlightened self-interest, more effective policies are required because the current ones are insufficient. A global, multidisciplinary effort is needed to slow the development of antibiotic resistance — it will take more than one shepherd to prevent our commons from being overgrazed. ■

EU COUNTRIES CUTTING BACK ON ANTIBIOTICS IN FARM ANIMALS

The amount of antibiotics used — adjusted here for different population sizes — varies widely from one EU country to another. Most importantly, every country on this list, except Spain, reduced its agricultural use of antibiotics from 2010 to 2011.



Megan Cully *is a freelance writer based in London.*

1. Zhu, Y.-G. *et al. Proc. Natl Acad. Sci. USA* **110**, 3435–3440 (2013).
2. European Medicines Agency. *Sales of Veterinary Antimicrobial Agents in 25 EU/EEA Countries in 2011* (2013).
3. Aarestrup, F. M. *et al. Am. J. Vet. Res.* **71**, 726–733 (2010).
4. Holis, A. & Ahmed, Z. *N. Engl. J. Med.* **369**, 2474–2476 (2013).

doi:10.1038/nature12212

Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome

Sheetal R. Modi[1], Henry H. Lee[1]†, Catherine S. Spina[1,2,3] & James J. Collins[1,2,3]

The mammalian gut ecosystem has considerable influence on host physiology[1–4], but the mechanisms that sustain this complex environment in the face of different stresses remain obscure. Perturbations to the gut ecosystem, such as through antibiotic treatment or diet, are at present interpreted at the level of bacterial phylogeny[5–7]. Less is known about the contributions of the abundant population of phages to this ecological network. Here we explore the phageome as a potential genetic reservoir for bacterial adaptation by sequencing murine faecal phage populations following antibiotic perturbation. We show that antibiotic treatment leads to the enrichment of phage-encoded genes that confer resistance via disparate mechanisms to the administered drug, as well as genes that confer resistance to antibiotics unrelated to the administered drug, and we demonstrate experimentally that phages from treated mice provide aerobically cultured naive microbiota with increased resistance. Systems-wide analyses uncovered post-treatment phage-encoded processes related to host colonization and growth adaptation, indicating that the phageome becomes broadly enriched for functionally beneficial genes under stress-related conditions. We also show that antibiotic treatment expands the interactions between phage and bacterial species, leading to a more highly connected phage–bacterial network for gene exchange. Our work implicates the phageome in the emergence of multidrug resistance, and indicates that the adaptive capacity of the phageome may represent a community-based mechanism for protecting the gut microflora, preserving its functional robustness during antibiotic stress.

Antibiotic treatment, an important and often necessary therapeutic intervention, can negatively affect the mammalian gut environment, potentially giving rise to immune[2] and metabolic deficiencies[8]. Studies on the disruption of intestinal homeostasis have focused on the resulting alterations in microbial composition[6,7]. However, investigation of the gut ecosystem has uncovered a myriad of resident phages[9], and it remains unclear how perturbation of the gut environment affects these symbionts. Phages can contribute genes that are advantageous to their microbial hosts[10,11], in turn promoting their own survival and propagation[12]. This gene flow suggests that phages may have an important role in the adaptation of the microbiome to stressful environments. We used a comparative metagenomic approach to explore the effects of antibiotic perturbation on functions encoded in the phageome, as well as to examine how antibiotic treatment alters the phage–bacterial ecological network.

We treated groups of young adult mice ($n = 5$) orally with physiologically relevant concentrations of ciprofloxacin (a quinolone that inhibits DNA synthesis) or ampicillin (a β-lactam that inhibits cell-wall synthesis), each with a respective control. We obtained collective faecal samples from each group after 8 weeks of treatment and purified phages as previously described[9,13]. DNA was extracted from phages and whole-genome amplified before performing shotgun 454 GS FLX+ pyrosequencing. We obtained a total of 440,792 quality reads, with a median read length of 477 nucleotides (210 megabases in total; Supplementary Fig. 1). Evaluation of contamination by quantification of bacterial 16S ribosomal RNA genes indicated that contaminating bacterial sequences represented less than 0.1% of our data, which was subsequently accounted for in all statistical analyses (see Supplementary Discussion and Supplementary Fig. 2).

Phage DNA sequences were compared to the non-redundant National Center for Biotechnology Information (NCBI) protein and environmental protein databases (BLASTX; E value $<10^{-5}$). Approximately 70% of reads were not assigned to previously sequenced genes (Supplementary Fig. 3), suggesting that the mouse phageome, like many other viral communities[14,15], harbours uncharacterized genetic material. We used the most significant BLAST alignment of a sequence, when available, to determine its phylogenetic origin. Most of the identifiable phages in our mouse phageomes (Supplementary Fig. 4a) were from the Caudovirales order, comprising the tailed phage families Siphoviridae, Podoviridae and Myoviridae, many of which are known to have a temperate life cycle. Because phage genomes incorporate bacterial genes, we also identified bacterial taxa; we found that 97% of phage-encoded bacterial genes were attributable to the four phyla known to dominate the gut (Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria; Supplementary Fig. 4b), consistent with the known hosts of the phages we detected.

We wondered whether antibiotic treatment leads to increases in phage-encoded genes for drug resistance. To investigate this, we compared DNA sequences in the phageome to an assembled database of antibiotic-resistance proteins (BLASTX; E value $<10^{-3}$, see Methods). We found that reads annotated as antibiotic-resistance genes were highly enriched in phage metagenomes from mice treated with ciprofloxacin or ampicillin compared with those from control mice (Z score $= 7.3$ and $Z = 7.0$, respectively; Supplementary Fig. 5; read annotations enumerated in Supplementary Table 1). We catalogued the resistance reservoir by annotating phage-encoded genes based on the drug class to which they confer resistance (Fig. 1a). Our analysis revealed that resistance to the administered drug class was enriched in phage metagenomes from antibiotic-treated mice, such that resistance to DNA-synthesis inhibitors was enriched in ciprofloxacin treatment ($Z = 2.6$), and resistance to cell-wall-synthesis inhibitors was enriched in ampicillin treatment ($Z = 5.0$). Additionally, upon drug treatment, new resistance genes were found in the phageome. For example, phages from ciprofloxacin-treated mice carried genes encoding numerous quinolone efflux pumps (for example, *norM*, *mexD*, *mexF*), and phages from ampicillin-treated mice carried genes encoding sensor and response regulators of cell-wall-synthesis inhibitors (for example, *vanRS*).

Of note, resistance to other, orthogonal drug classes was also overrepresented in the antibiotic-perturbed phageomes. Both treatments led to significant enrichment of resistance to antibiotics that target protein synthesis, and ciprofloxacin treatment also led to significant enrichment of resistance to cell-wall synthesis inhibitors (Fig. 1a). This cross-resistance was mediated by drug-specific inactivators (for example, chloramphenicol acetyltransferases), as well as multidrug-resistance exporters (for example, *mdtK*; Supplementary Table 1). Together, these

[1]Howard Hughes Medical Institute, Department of Biomedical Engineering, and Center for BioDynamics, Boston University, Boston, Massachusetts 02215, USA. [2]Boston University School of Medicine, 715 Albany Street, Boston, Massachusetts 02118, USA. [3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02118, USA. †Present address: Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

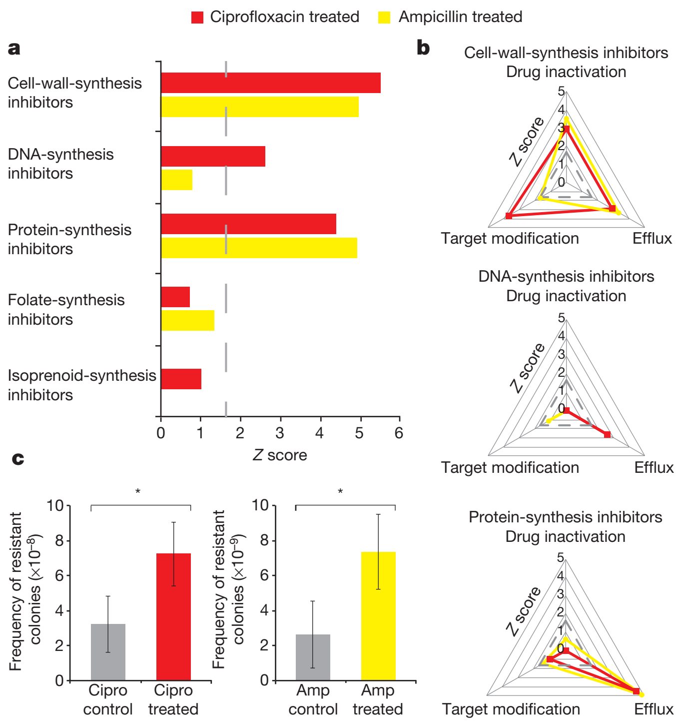


Figure 1 | Antibiotic resistance is enriched in phage metagenomes following drug perturbation in mice. **a**, **b**, Z scores are shown for sequencing reads annotated as antibiotic-resistance genes in phages from ciprofloxacin-treated (red) and ampicillin-treated (yellow) mice in comparison with respective control mice. Dashed lines correspond to a Z score of 1.65 ($P = 0.05$). Phage-encoded resistance genes were classified according to the drug class to which they confer resistance (**a**) and by their mechanism of resistance (**b**). **c**, Frequency of colonies resistant to ciprofloxacin (Cipro; 1 μg ml^{-1}) upon infection of microbiota with phages from ciprofloxacin-treated mice or phages from control mice (left), and frequency of colonies resistant to ampicillin (Amp; 4 μg ml^{-1}) upon infection of microbiota with phages from ampicillin-treated mice or phages from control mice (right). P values from Mann–Whitney U-test; $n > 12$. Data show mean ± standard error of the mean (s.e.m.). *$P < 0.05$.

findings implicate the phage metagenome as a potential source of multidrug resistance during antibiotic treatment of the host.

We aimed to understand the specific mechanisms represented in phage-encoded genes conferring resistance to the most significantly enriched classes of antibiotics: inhibitors of cell-wall synthesis, DNA synthesis, and protein synthesis. We categorized resistance genes according to primary resistance mechanisms, which include modification or protection of the drug target (target modification), enzymatic inactivation of the drug (drug inactivation), and transport of the drug out of the cell (efflux)[16]. Using this framework to classify phage-encoded resistance genes, our analysis revealed that antibiotic treatment led to disparate resistance mechanism profiles for each drug class (Fig. 1b). Analysis of resistance to cell-wall-synthesis inhibitors showed that all types of resistance mechanisms were significantly enriched with both ciprofloxacin treatment and ampicillin treatment. By contrast, resistance to DNA-synthesis and protein-synthesis inhibitors occurred predominantly by efflux. Resistance to protein-synthesis inhibitors occurred by target modification and drug inactivation mechanisms at low levels and, in accordance with its rarity, resistance to DNA-synthesis inhibitors through drug inactivation was not detected. These data probably reflect resistance mechanisms that are both environmentally available in the gut ecosystem and impose sustainable *in vivo* fitness costs. As continued treatment with an antibiotic invariably leads to its own resistance, mechanisms that enable cross-resistance encoded by the phageome may be an important consideration when selecting subsequent therapeutics.

We next sought to demonstrate that phages from antibiotic-treated mice confer increased drug resistance to the host-associated bacterial community. We assessed the frequency of resistant isolates from aerobically cultured naive microbiota that were infected *ex vivo* with phages from antibiotic-treated or control (untreated) mice. Our results show that this fraction of microbiota infected with phages from mice administered ciprofloxacin or ampicillin yielded two to three times more colonies resistant to the respective drug than the aerobically cultured fraction of microbiota infected with phages from control mice (Fig. 1c). These data indicate that phages from antibiotic-treated mice can contribute relevant functional advantages to their microbial hosts.

We next took a systems-level approach and classified other phage-encoded genes into functional pathways described by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We depicted enriched functional changes in phage metagenomes after antibiotic treatment as a network diagram (Fig. 2; abundances shown in Supplementary Fig. 6). Among the most significantly enriched pathways were functional properties related to the mode of action of the administered drug (Supplementary Table 2). Phageomes from ampicillin-treated mice were enriched for the amino sugar and nucleotide metabolism pathway (part of the broader carbohydrate metabolism process; $Z = 5.6$), indicating over-representation of genes related to synthesis of cell-wall constituents, and increases in these components have been shown to be requisite for drug resistance in clinical isolates[17]. Additionally, we found that phageomes from ciprofloxacin-treated mice were enriched for replication- and repair-related pathways, including base excision repair ($Z = 6.1$), nucleotide excision repair ($Z = 7.4$) and homologous recombination ($Z = 11.2$). Included in these pathways are members of the GO system for the repair of DNA oxidative lesions, which has been demonstrated to reduce cytotoxicity due to a range of antibiotic classes[18]. Also represented are members of the DNA-damage-inducible SOS system, known to provide protection against antibiotic-mediated cell death and induce the development of resistance-conferring mutations[19]. Furthermore, hyper-recombination has been shown to promote multidrug resistance phenotypes[20]. These results show that under drug treatment, the phageome encodes diverse mechanisms for modulating antibiotic susceptibility.

We also observed that phage metagenomes from antibiotic-treated mice were enriched for microbial functions that contribute to host metabolism (Fig. 2 and Supplementary Table 2). Phageomes from ciprofloxacin-treated mice were uniquely enriched for pathways relevant to the metabolism of cofactors and vitamins, including thiamine, an essential nutrient provided by the microbiome. Microbiota ferment polysaccharides indigestible by the host alone; metabolism of these sugars enables bacterial survival in and colonization of the gut environment[21] and, as a beneficial consequence, provides energy to the host[4]. We found that polysaccharide-degradation genes, specifically related to metabolism of starch, cellulose, lactose and fructans (plant-derived fructose polymers), were enriched with antibiotic treatment (Fig. 3a). Genes coding for carbohydrate active enzymes (CAZymes), which enable bacteria to ferment a variety of dietary- and host-sourced glycans, were also enriched with antibiotic treatment and were represented by a range of glycoside hydrolase and glycosyltransferase families (Supplementary Fig. 7 and Supplementary Table 3). Because many gut microbes express only a specific array of carbohydrate-degrading enzymes[22], bacteria that acquire these genes from the phage reservoir may gain additional foraging capacity and, consequently, a selective growth advantage. These results suggest that the phageome may be an adaptive repository for functions important for the host–commensal relationship that may otherwise be depleted by antibiotic perturbation. (See Supplementary Information for additional discussion.)

We next aimed to elucidate the phylogenetic basis of these phage-encoded bacterial functions. In Fig. 3b, we illustrate the taxonomic composition of all sequences of bacterial origin and sequences annotated with enriched functions following drug perturbation. Examining bacterial phylotypes that contribute antibiotic-resistance genes to the phage metagenome, we found a comparatively high representation of the Clostridia class and a low representation of the Bacilli class.

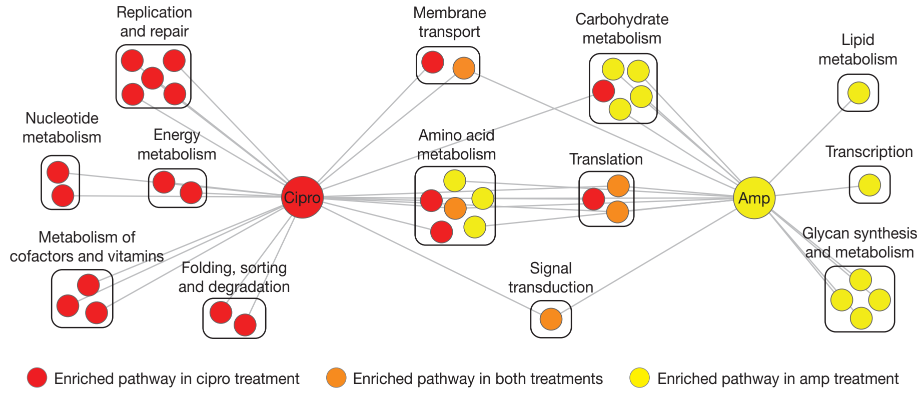


Figure 2 | Broad bacterial functions are enriched in phage metagenomes following drug perturbation in mice. Network depicts KEGG pathways significantly enriched under antibiotic treatment compared with controls. Treatments are represented by large nodes; enriched pathways are represented by small nodes, grouped by their higher-level processes and coloured by the treatment condition (red, ciprofloxacin; yellow, ampicillin; orange, common to both treatments). In total, we identified 24 out of 188 pathways that were enriched with ciprofloxacin treatment ($Z \geq 3.46$, Bonferroni corrected), and 18 out of 178 pathways that were enriched with ampicillin treatment ($Z \geq 3.43$, Bonferroni corrected). Amp, ampicillin; cipro, ciprofloxacin.

Notably, a large fraction of CAZyme-annotated sequences originated from the Bacteroidetes class, which comprises members found to have diverse capabilities for carbohydrate metabolism[22]. Investigation of thiamine metabolism reveals that the Bacilli and Verrucomicrobia classes constitute a large proportion of these annotations. As the phageome reflects emergent properties of its environment[15], phylogenetic analyses of phage-encoded elements may more broadly enable the identification of bacteria actively contributing to specific functions in the gut environment.

Our results show that the phageome harbours a diversity of potentially beneficial functional elements in the face of antibiotic perturbation. However, the extent to which this genetic reservoir is accessible to members of the microbiota remains unclear[23]. To investigate this, we sought to elucidate the phage–bacterial ecological network and how it changes under antibiotic treatment. We approximated the network of phage–microbe interactions with relationships identified through the reconstruction and analysis of individual viral genomes. *De novo* assembly was accomplished using stringent parameters (see Methods). Reconstructed viral genomes are composed of a mosaic of bacterial genes, and we used the phylogenetic origins of these sequences to determine putative phage–bacterial associations (Fig. 4). Our resulting network recapitulated known interactions, including the lysogenic relationships of foodborne pathogens, such as bacteriophage φ3626 infection of *Clostridium perfringens*[24] and Siphoviridae Listeria phage A500 infection of *Listeria monocytogenes*[25]. Importantly, antibiotic treatment leads to widespread restructuring of the phage–bacterial ecological network (Fig. 4). These data show that new links between phages and bacteria are formed with drug treatment, giving rise to significantly greater network connectivity (Supplementary Fig. 8). This increased connectivity is reflective of phages broadly, as more bacterial species are associated with a given phage (Supplementary Fig. 9). These results suggest that antibiotic treatment increases the frequency of phage integration and stimulates broad host range, which promotes a functional reservoir that is both genetically diverse and highly accessible to gut bacteria.

Although the phageome is a highly connected network for gene exchange, the functional consequence of acquiring genetic material from this reservoir depends on the molecular context of the host bacterium. Acquisition of a single gene may enhance an existing function by gene dosage or enable a novel phenotype. As some proteins rely on additional machinery, subsequent horizontal gene-transfer events may be required to produce a phenotype. Moreover, redundantly acquired genes may gain new functionality through paralogueous evolution.

We demonstrate that antibiotic treatment enriches the phage metagenome for stress-specific and niche-specific functions, while mediating changes in the topology of the phage–bacterial ecological network to potentiate accessibility of these genetic elements. Functional resilience of the microbiome following environmental perturbation has been empirically documented and has engendered interest in the restorative forces that return the commensal flora to its pre-perturbed state[6,7,26]. Our results implicate phage encapsulation of adaptive signatures as a community-based mechanism for functional robustness in the gut environment during stress. Of note, antibiotic treatment can also prime the gut environment for pathogen invasion[2], and our findings have potential implications for the emergence of drug resistance and evasion strategies in pathogenic populations. Cohabitation of phages and bacteria in the gut ecosystem is probably governed by complex and dynamic interactions, particularly during stress perturbation. Additional work is needed to discern the selective pressures imposed on each member of this community and the resulting mechanisms that influence the

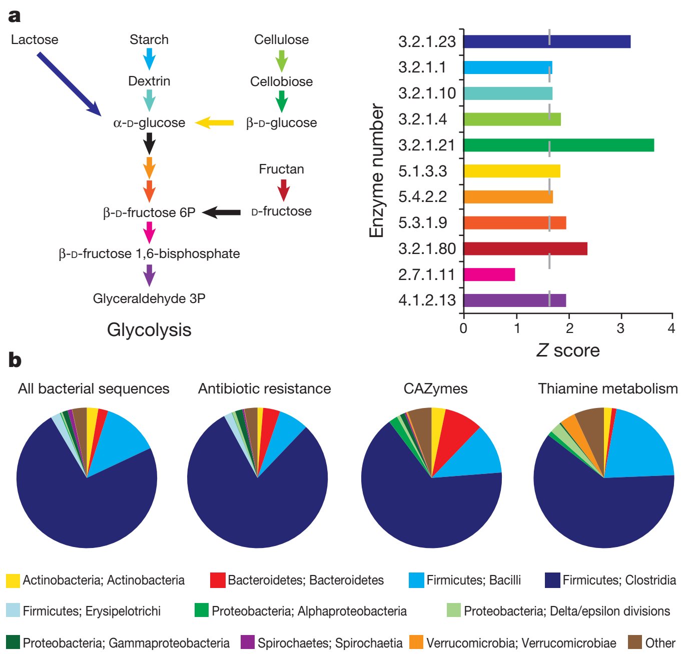


Figure 3 | Investigation of bacterial functions encoded in phages. **a**, Bacterial enzymes from sugar metabolism to glycolysis (left) with corresponding Z scores in phages from drug-treated mice in comparison with control mice (right). Dashed line corresponds to a Z score of 1.65 ($P = 0.05$). **b**, Class-level taxonomic distribution of all sequences of bacterial origin identified in phage sequencing data (far left) and sequences annotated with enriched functions following drug perturbation. 'Other' constitutes taxa that contributed less than 1% to all distributions.

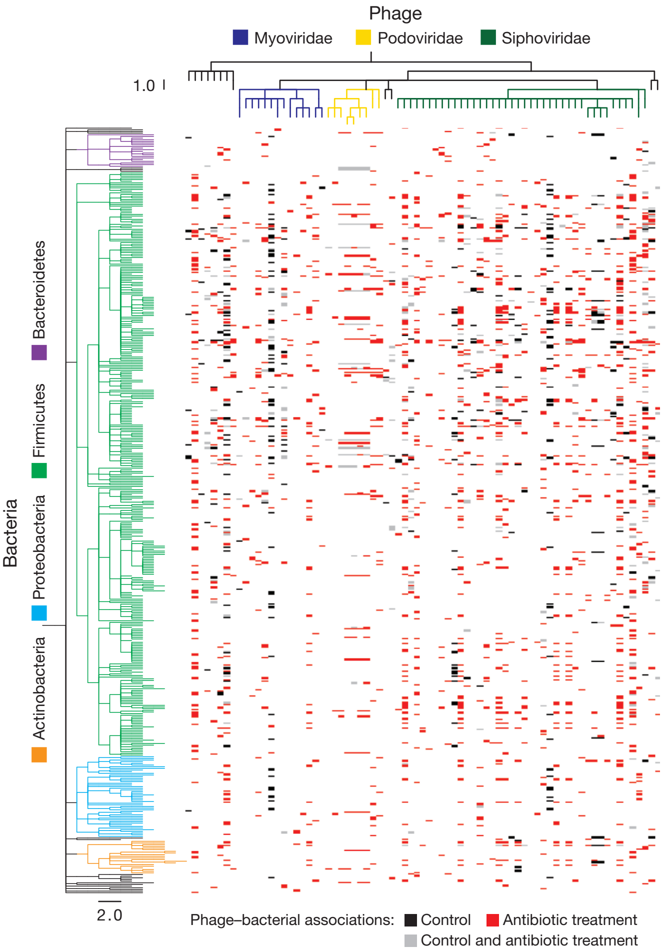


Figure 4 | The phage–bacterial ecological network. Dashes represent associations between virotypes and bacterial species identified from phylogenetic analysis of reconstructed phage genomes. Phage–bacterial associations only in control metagenomes (black), only in drug-treated metagenomes (red), and commonly identified in control and drug-treated metagenomes (grey). Data are the union of associations identified in 50 assemblies of randomly sampled reads from each treatment.

encoding and progressive enrichment of functionally beneficial genes in the phageome. Phage-mediated gene flow may be an important phenotypic buffer for bacterial communities, and further investigation of the adaptive reservoir of the phageome and the dynamic nature of the phage–bacterial ecological network may prove critical to understanding the influence of the gut ecosystem on host physiology.

METHODS SUMMARY

Groups of 6-week-old female FVB mice were treated with antibiotics in their drinking water to achieve doses of 28.5 mg kg^{-1} day^{-1} ampicillin or 12.5 mg kg^{-1} day^{-1} ciprofloxacin. Control groups were supplied with standard drinking water (ampicillin) or alkaline water (ciprofloxacin). After 8 weeks, we harvested collective faecal samples from each group. Viral purification from faecal samples was performed as previously described[9,13]. DNA was extracted from each viral sample and whole-genome amplified in three separate reactions. Equimolar concentrations of multiplexed samples were pooled on a single plate for 454 GS FLX+ pyrosequencing. To analyse sequencing reads, an antibiotic resistance database was assembled using sequences from the Antibiotic Resistance Genes Database (ARDB)[27] and UniProt proteins were annotated with the Gene Ontology term "antibiotic breakdown". Custom perl scripts were written to annotate sequences with KEGG (v.61.0). Enrichment between treatment and control was calculated by random sampling with replacement ($n = 10{,}000$), and Z scores were computed from the resulting normal distribution. Contigs were assembled using the Roche 454 GS *De novo* Assembler with default parameters, except for a minimum overlap of 100 bp and a minimum identity of 100%.

Full Methods and any associated references are available in the online version of the paper.

Received 3 July 2012; accepted 22 April 2013.
Published online 9 June 2013.

1. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331,** 337–341 (2011).
2. Brandl, K. *et al.* Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. *Nature* **455,** 804–807 (2008).
3. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480,** 241–244 (2011).
4. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
5. Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* **333,** 101–104 (2011).
6. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6,** e280 (2008).
7. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108** (Suppl. 1), 4554–4561 (2011).
8. Antunes, L. C. *et al.* Effect of antibiotic treatment on the intestinal metabolome. *Antimicrob. Agents Chemother.* **55,** 1494–1503 (2011).
9. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466,** 334–338 (2010).
10. Oliver, K. M., Degnan, P. H., Hunter, M. S. & Moran, N. A. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* **325,** 992–994 (2009).
11. Chen, J. & Novick, R. P. Phage-mediated intergeneric transfer of toxin genes. *Science* **323,** 139–141 (2009).
12. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438,** 86–89 (2005).
13. Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L. & Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nature Protocols* **4,** 470–483 (2009).
14. Breitbart, M. *et al.* Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185,** 6220–6223 (2003).
15. Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452,** 629–632 (2008).
16. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406,** 775–781 (2000).
17. Hanaki, H. *et al.* Activated cell-wall synthesis is associated with vancomycin resistance in methicillin-resistant *Staphylococcus aureus* clinical strains Mu3 and Mu50. *J. Antimicrob. Chemother.* **42,** 199–209 (1998).
18. Foti, J. J., Devadoss, B., Winkler, J. A., Collins, J. J. & Walker, G. C. Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science* **336,** 315–319 (2012).
19. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nature Rev. Microbiol.* **8,** 423–435 (2010).
20. Hanage, W. P., Fraser, C., Tang, J., Connor, T. R. & Corander, J. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science* **324,** 1454–1457 (2009).
21. Chang, D. E. *et al.* Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc. Natl Acad. Sci. USA* **101,** 7427–7432 (2004).
22. Xu, J. & Gordon, J. I. Honor thy symbionts. *Proc. Natl Acad. Sci. USA* **100,** 10452–10459 (2003).
23. Stern, A., Mick, E., Tirosh, I., Sagy, O. & Sorek, R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* **22,** 1984–1994 (2012).
24. Zimmer, M., Scherer, S. & Loessner, M. J. Genomic analysis of *Clostridium perfringens* bacteriophage φ3626, which integrates into *guaA* and possibly affects sporulation. *J. Bacteriol.* **184,** 4359–4368 (2002).
25. Dorscht, J. *et al.* Comparative genome analysis of *Listeria* bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. *J. Bacteriol.* **191,** 7206–7215 (2009).
26. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220–230 (2012).
27. Liu, B. & Pop, M. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **37,** D443–D447 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Bodi and J. Schiemer at the Tufts Genomic Core for their sequencing assistance and A. Green and K. Pardee for ultracentrifugation help. This work was supported by the Howard Hughes Medical Institute and the National Institutes of Health Director's Pioneer Award Program.

Author Contributions All authors designed the study. C.S.S. oversaw the mouse work. S.R.M. and H.H.L. performed and analysed the experiments, with conceptual insight provided by J.J.C. S.R.M., H.H.L. and J.J.C. prepared the manuscript.

Author Information Virome data sets have been deposited in the NCBI Sequence Read Archive under accession SRP021521. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.J.C. (jcollins@bu.edu).

METHODS

Mouse study. All experiments involving animals were pre-reviewed and approved by the Boston Children's Hospital Institutional Animal Care and Use Committee. Groups of separately housed 6-week-old female FVB/NJ mice ($n = 5$; Jackson Laboratory) were treated with ampicillin (142.5 mg l^{-1}) or ciprofloxacin (62.5 mg l^{-1}) in their drinking water. The corresponding dosage was 28.5 mg kg^{-1} day^{-1} ampicillin and 12.5 mg kg^{-1} day^{-1} ciprofloxacin, based on the average mouse weight of 20 g and an approximate intake of 4 ml per day. Control mouse groups were supplied with standard drinking water (ampicillin) or alkaline water (ciprofloxacin). Treatments were refreshed twice per week. Mice were housed in sterile conditions and received autoclaved chow during the course of this study. We harvested fresh collective faecal samples from each group to obtain ample material (3–4 g) for purification. Samples were stored at −80 °C before use.

Viral purification and preparation of genomic DNA. Viral purification from faecal samples of mice after 8 weeks was performed as previously described[9,13]. An aliquot of the viral preparation (1.5 g ml^{-1} layer from ultracentrifugation) was stained with SYBR gold and visualized with epifluorescence microscopy to verify the absence of bacterial contamination. Viral particles were concentrated and desalted using an Ultra-4 Centrifugal Filter Unit (Ultracel-30K MWCO; Millipore) to a volume of ~200 µl. Concentrated viral samples were treated with DNase (0.2 mg ml^{-1}) and samples were passed through a 0.22 µM filter to ensure no procedural contamination was introduced. Genomic DNA was extracted using the QiaAMP DNA mini kit (Qiagen) as per the protocol for viral DNA detailed in the manual. Genomic DNA was amplified using the Illustra Genomiphi v2 kit (GE) according to the manufacturer's instructions. For each sample, we pooled amplified DNA from three separate reactions to minimize bias.

Next-generation sequencing. Viral DNA was submitted to the Tufts Genomic Core for library preparation and shotgun sequencing. Equimolar concentrations of multiplexed (Rapid Library MID) samples were pooled on a single plate and pyrosequenced using the 454 GS FLX+ platform. Resulting sequences were filtered by removing duplicates using the tool available at http://microbiomes.msu.edu/replicates/[28] with the following parameters: sequence identity cut-off = 97%; length difference requirement = 0; number of beginning base pairs to check = 20.

Antibiotic-resistance annotations. To facilitate annotation of antibiotic-resistance genes, we assembled a database consisting of the ARDB[27] and UniProt proteins annotated with the Gene Ontology (GO) term "antibiotic breakdown" (GO: 0017001) to achieve a total of 12,687 protein sequences. The functional annotations of the proteins in this database are supported by either experimental validation or high-quality computational prediction. Viral DNA sequences were compared with this database using BLASTX, and sequences with an E value $<10^{-3}$ were deemed significant. This cut-off was selected to maintain a consistent stringency, by accounting for database size, with functional annotation to the NCBI databases.

Functional annotations. Phage DNA sequences were compared with the non-redundant NCBI protein (nr) and environmental protein (env_nr) databases (BLASTX; E value $<10^{-5}$). Sequences were annotated with KEGG[29] (v.61.0) using custom Perl scripts interfaced with the KEGG API. The most significant BLAST hit that resulted in a KEGG annotation was used, and we included all annotations in the event that a hit had multiple annotations. KEGG orthologue annotations were compared with glycoside hydrolases and glycosyltransferases found on http://www.cazy.org[30] to identify CAZyme-encoding genes.

Statistical testing. To compare the functional annotations of two metagenomic data sets, A and B, we generated a distribution for each data set, A and B, reflecting the number of annotations from 10,000 trials of random sampling with replacement, sampled at the number of reads in the comparison data set. To account for the effects of contaminating bacterial DNA in our comparative metagenomic analyses, we assumed that contamination would be uniformly distributed and therefore we randomly discarded bacterial reads in each sampling trial according to the amount of contamination detected by quantitative PCR. We compared the number of annotations identified in a given sample to the comparison distribution and determined a Z score, calculated as $(x-\mu)/\sigma$, where, for example, x is the raw number of annotations in sample A, μ is the mean number of annotations in the distribution for sample B, and σ is the standard deviation of B's distribution. In essence, this results in two Z scores, one comparing A to B and one comparing B to A, and consequently the minimum $|Z|$ was used. According to the central limit theorem, random sampling with replacement results in a normal distribution, so Z scores >1.65 ($P < 0.05$) were considered enriched. In KEGG pathway analysis, Bonferroni was used to correct for multiple hypotheses, where the P-value cut-off $0.05/n$ (n was the number of total third-level pathways identified in our phage metagenomes) was converted to a Z score.

Phage infection of microbiota. Microbiota were isolated from faecal samples of naive mice as previously described[31], except PBS plus 0.1% cysteine supernatants were plated on four separate Luria–Bertani (LB) agar plates. Colonies were grown aerobically for 24 h before plates were scraped with 2 ml LB, amassed, and cultured at 37 °C and 300 r.p.m. for 2 h. This mixture was stored as 150 µl aliquots in 15% glycerol at −80 °C. For each experiment, an entire aliquot was used as the inoculum to minimize growth biases. Phages were isolated as described above. Owing to limited sample availability, we harvested phages from mice that had been treated for 5 weeks for ciprofloxacin experiments and phages from mice that had been treated for 3 weeks for ampicillin experiments, along with phages from control mice, respectively. Phage preparations from drug-treated mice and control mice were diluted to the same volume, split equivalently, and incubated with 0.25 ml microbiota (cultured to exponential phase in LB plus 0.2% maltose) with 5 mM $CaCl_2$ and 10 mM $MgSO_4$. Phage–microbiota mixtures were allowed to adsorb for one hour at 37 °C (no shaking). Phage-infected microbiota were then pelleted, resuspended in fresh LB, and plated on LB agar plates with ciprofloxacin (1 µg ml^{-1}) or LB agar plates with ampicillin (4 µg ml^{-1}). A 10 µl aliquot was serially diluted and plated onto no-drug LB agar plates. Frequency was calculated as: number of colonies on drug plate divided by number of colonies on no-drug plate. Additionally, the basal frequency of resistant isolates from microbiota was measured as described in the previous sentence in the absence of phages, and this frequency was confirmed to be lower than that from microbiota infected with phages from either treated or untreated mice.

Quantification of 16S rRNA. Quantitative (q)PCR was used to measure levels of the 16S rRNA gene in our viral preparations. We used the universal primers 8F and 338R[9], and qPCR was performed using the SYBR Green I Master kit and the LightCycler 480 (Roche) according to the manufacturer's instructions.

Contig assembly and identification of phage–bacterial associations. Contigs were assembled using the Roche 454 GS *De novo* Assembler with default parameters, except for a minimum overlap of 100 bp and a minimum identity of 100% to minimize erroneous alignments. Phage–bacterial associations were determined by computing the combination of phage phylogenetic annotations and bacterial phylogenetic annotations on a given contig, and non-redundant phage–bacterial associations were amassed for all contigs. To evenly compare the number of phage–bacterial associations across samples, we performed 50 assemblies using 60,000 randomly selected sequences with replacement from each sample. The data presented in Fig. 4 are representative of these 50 assemblies such that an association was illustrated if it was present in at least one assembly analysis. We computed the mean number of phage–bacterial associations for a given sample (Supplementary Fig. 8) from the number of non-redundant associations found in each assembly trial. Significance was determined by comparing sample values using the Mann–Whitney U-test. We computed the mean bacteria-to-phage ratio for a given sample (Supplementary Fig. 9) from the number of bacterial species associated with a given phage for all phages using the union of associations from 50 assemblies shown in Fig. 4. Significance was determined by comparing sample values using the Mann–Whitney U-test.

28. Gomez-Alvarez, V., Teal, T. K. & Schmidt, T. M. Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* **3,** 1314–1317 (2009).
29. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
30. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37,** D233–D238 (2009).
31. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108,** 6252–6257 (2011).

The Journal of Antibiotics (2013) 66, 709–712

www.nature.com/ja

ORIGINAL ARTICLE

Comparative study of the mutant prevention concentrations of vancomycin alone and in combination with levofloxacin, rifampicin and fosfomycin against methicillin-resistant *Staphylococcus epidermidis*

Li-guang Liu[1,4], Yu-lin Zhu[1,4], Li-fen Hu[1], Jun Cheng[1], Ying Ye[1,2,3] and Jia-bin Li[1,2,3]

No mutant-prevention concentration (MPC) with methicillin-resistant *Staphylococcus epidermidis* (MRSE) has been reported. The study aimed to evaluate the propensity of vancomycin individually and in combination to prevent MRSE from mutation. A total of 10 MRSE clinical isolates were included in the study. Susceptibility testing demonstrated that the susceptibility rates to vancomycin, rifampicin, levofloxacin and fosfomycin were 100, 100, 50 and 90%, respectively. The fractional inhibition concentration indices (FICI) for vancomycin combined with rifampicin, levofloxacin or fosfomycin were $\geqslant$1.5 but $\leqslant$2, $\geqslant$1.5 but $\leqslant$2, and $>$0.5 but $\leqslant$1.5, respectively, implying indifferent interactivity. The MPC with susceptible strains was determined to be the lowest antibiotic concentration inhibiting visible growth among 10^{10} CFU on four agar plates (9 cm in diameter) after a 72-h incubation at 37 °C. The MPCs were 16$\sim$32, $>$64, $\geqslant$64 and 4$\sim$16 μg ml^{-1} for vancomycin, rifampicin, fosfomycin and levofloxacin, respectively. The vancomycin MPCs of combinations with fosfomycin (32 μg ml^{-1}), levofloxacin (2 μg ml^{-1}) and rifampicin (2 or 4 μg ml^{-1}) were 1$\sim$4, 16$\sim$32 and 16$\sim$32 μg ml^{-1}, respectively. Against mutants selected by vancomycin, rifampicin, levofloxacin and fosfomycin individually, antibiotics had standard MICs of 2$\sim$4 μg ml^{-1} for vancomycin, $>$64 μg ml^{-1} for rifampicin, 4$\sim$8 μg ml^{-1} for levofloxacin and $\geqslant$64 μg ml^{-1} for fosfomycin. Thus single-step mutation can lead to resistance of MRSE to rifampicin, levofloxacin and fosfomycin, rather than non-susceptibility to vancomycin. Vancomycin–fosfomycin combination might be a superior alternative to vancomycin in blocking the growth of MRSE mutants, especially for high-organism-burden infections.
The Journal of Antibiotics (2013) **66,** 709–712; doi:10.1038/ja.2013.87; published onine 28 August 2013

Keywords: combination therapy; drug resistance; mutant prevention concentration; MRSE

INTRODUCTION

Staphylococcus epidermidis has become a prevalent cause of nosocomial infections due to increasing use of broad-spectrum antimicrobials, immunosuppressants and biomaterials in the clinic, and methicillin-resistant *S. epidermidis* (MRSE) has accounted for a majority of *S. epidermidis* nosocomial infections.[1,2] To date, vancomycin still remains the drug of choice against MRSE infections. However, decreased vancomycin susceptibility in staphylococci (including heterogeneous and homogeneous vancomycin-intermediate staphylococci) has been reported all over the world. Though there are alternatives to vancomycin such as linezolid, daptomycin, tigecycline and quinupristin-dalfopristin, they are expensive and not easy to access, especially in developing countries. With a limited pool of available antimicrobial agents capable of treating MRSE infections, the suppression of further emergence of resistance is of vital importance.

The 'mutant selection window' hypothesis[3,4] proposes a way to fight against the drug resistance from a new viewpoint. It postulates that maintaining antimicrobial concentration at infection sites within mutant selection window, a range between MIC and mutant prevention concentration (MPC) is expected to select for and enrich resistant mutants. MPC is defined as the minimal concentration required to inhibit the growth of the least susceptible single-step mutant and approximated experimentally as the lowest concentration that allows no colony growth when more than 10^{10} cells are applied to drug-containing culture medium. Maintaining antimicrobial concentration at infection sites above MPC or concurrent use of

[1]Department of Infectious Diseases, the First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China; [2]Anhui Center for Surveillance of Bacterial Resistance, Hefei, Anhui, China and [3]Institute of Bacterium Resistance, Anhui Medical University, Hefei, Anhui, China
[4]These authors contributed equally to this work.
Correspondence: Dr Y Ye, Department of Infectious Diseases, the First Affiliated Hospital of Anhui Medical University, Jixi road 218, Hefei, Anhui 230022, China.
E-mail: yeying2@139.com
Received 9 June 2012; revised 13 December 2012; accepted 8 May 2013; published online 28 August 2013

two antimicrobials at concentrations above MIC will block the amplification of mutants and final resistance because bacteria will have to acquire two concurrent mutations for growth, which rarely exists *in vivo*. Because MPC determinations usually apply an inoculum size of approximately 10^{10} CFU containing potentially resistant mutant subpopulations, MPC is a susceptibility parameter that can measure better the relative potential for selection of resistant mutants by antimicrobials than traditional susceptibility testing.

Ideally, MPC should be determined for organism–antibiotic pairings for which spontaneous mutation is a common mechanism of resistance. Available data indicate that the resistance of staphylococci to vancomycin,[5–7] levofloxacin,[8] rifampicin[9] and fosfomycin[10] comply with the above mentioned principle. So far, no study of the MPC for MRSE has been published yet. This study aimed mainly to compare the mutant-prevention efficacies of vancomycin used alone and in combination with levofloxacin, rifampicin or fosfomycin. Rifampicin, levofloxacin and fosfomycin were chosen in this study because they are cheap and penetrate well where vancomycin penetrates poorly (for example, the lung, cerebrospinal fluid and bone joint). In addition, the susceptibility rates of *S. epidermidis* to fosfomycin[11] and rifampicin[12,13] remain fairly high.

MATERIALS AND METHODS

Strains and drugs

Totally 10 MRSE strains isolated from different clinical specimens (7 blood, 1 wound swabs, 2 catheter tips) were chosen for the study. These strains were from different patients (5 neonates, 2 elders, 2 adults, 1 child) hospitalized in three general hospitals in Hefei city, Anhui province, China. Specific identification of MRSE was conducted previously using intragenic primer sets for staphylococcus 16SrRNA,[14] mecA gene sequences[14] and *S. epidermidis* species-specific gene sequence.[15] *Staphylococcus aureus* ATCC 29213 was stored in Anhui Center for Surveillance of Bacterial Resistance and used as the quality control strain in MIC determination and synergy testing. The antimicrobials in this study and their suppliers were as follows: levofloxacin (LVX), rifampicin (RIF) and fosfomycin (FOF) standards were obtained from the National Institute for the Control of Pharmaceutical and Biological Products (NICPBP; Beijing, China). Vancomycin (VAN) was purchased from Eli Lilly Japan K.K, Seishin Laboratories (Kobe, Japan). Glucose-6-phosphate was purchased from Sigma (Sigma, Deisenhofen, Germany).

MIC determination

The MICs were determined using the standard agar doubling dilution method. When MIC and MPC of fosfomycin or fosfomycin-containing combination were to be determined, Mueller–Hinton agar (MHA) should be supplemented with $25\,\mu g\,ml^{-1}$ glucose-6-phosphate.[16] ATCC 29213 (*S. aureus*) was included as the quality-control strain. An MRSE isolate was considered susceptible, intermediate or resistant for respective MIC values of $\leqslant 4$, 8–16 or $\geqslant 32\,\mu g\,ml^{-1}$ for vancomycin, $\leqslant 1$, 2 or $\geqslant 4\,\mu g\,ml^{-1}$ for levofloxacin, and $\leqslant 1$, 2 or $\geqslant 4\,\mu g\,ml^{-1}$ for rifampicin, according to the CLSI guidelines.[16] An MRSE isolate was considered susceptible or resistant for fosfomycin MIC values of $\leqslant 32$ or $>32\,\mu g\,ml^{-1}$, respectively, according to EUCAST guidelines[17] (the breakpoint concentrations of fosfomycin against staphylococci are not available in the CLSI guidelines).

Synergy testing

Checkerboard synergy testing was performed for all the susceptible strains with two-fold agar dilutions on MHA.[18] MHA was also supplemented with $25\,\mu g\,ml^{-1}$ glucose-6-phosphate for synergistic testing of the fosfomycin–vancomycin combination. The concentrations were set according to the MIC values of the preliminary susceptibility tests. To be exact, vancomycin concentrations (doubling dilutions) ranging $1/64\text{-}4 \times$ MIC in the columns and rifampicin, levofloxacin or fosfomycin concentrations (doubling dilutions) ranging $1/64\text{-}4 \times$ MIC along the rows were combined with each other on the agar plate in a checkerboard style. The MIC of individual drug was tested again in parallel along the row or column. Agar plates were inoculated using a multipoint inoculator (5×10^4 CFU per spot) and incubated at 37 °C for 24 h. The standard quality-control strain ATCC 29213 was incubated with each run. Fractional inhibitory concentration (FIC) was calculated as the MIC of a drug in combination divided by the MIC of the drug alone, and the FIC index (FICI) was obtained by adding the FICs. The results of FIC indices were interpreted as synergistic (FICI$\leqslant$0.5), indifferent (0.5$<$FICI$<$4) and antagonistic (FICI$\geqslant$4) according to the most relevant criteria.[19]

MPC determination and MIC retesting of putative mutants

MPCs were determined as described by Zhao *et al.*[20] Briefly, the tested bacteria were cultured in Mueller–Hinton broth and cultured for 24 h. The suspension was centrifuged (4000 *g* for 10 min) and re-suspended in Mueller–Hinton broth to yield a concentration of about 3×10^{10} CFU ml^{-1}. Finally, each agar plate containing known antibiotic concentrations was inoculated with about 3.0×10^9 CFU of tested bacteria (100 μl bacterial suspension), with four plates (9 cm in diameter) for a given antibiotic concentration. The MPC was determined to be the lowest antibacterial concentration that completely inhibited visible bacterial growth among 10^{10} CFU after 72 h incubation at 37 °C in ambient air. Different doubling dilution concentrations were prepared for each antibiotic–bacteria combination, with $1 \times$ MIC as the lowest concentration. The top concentrations prepared for MPC determinations were 64, 64 and $16\,\mu g\,ml^{-1}$ for vancomycin, rifampicin and levofloxacn, respectively, which were all above the attainable maximum serum concentrations at the currently recommended doses (the corresponding dosing regimens are 1000 mg i.v. twice daily for vancomycin, 450 or 600 mg po or i.v. once daily for rifampicin, and 400, 500 or 750 mg po or i.v. once daily for levofloxacin). The top concentration prepared for fosfomycin MPC determination was $64\,\mu g\,ml^{-1}$, which is double the resistant or susceptible breakpoint concentration of *Staphylococcus* to fosfomycin ($32\,\mu g\,ml^{-1}$) according to EUCAST.[17] The concentrations of levofloxacin ($2\,\mu g\,ml^{-1}$) and rifampicin (2 or $4\,\mu g\,ml^{-1}$) used in combination with vancomycin reflect their mean serum concentrations at steady state in healthy adults[21] and adult pulmonary tuberculosis patients,[22,23] respectively. The concentration of fosfomycin ($32\,\mu g\,ml^{-1}$) used in combination with vancomycin equals the susceptible breakpoint concentration of staphylococci but is far below the mean serum concentrations at steady state according to currently recommended dosing regimens.[24] It should be mentioned that MPCs were determined only for susceptible isolates. Independent cells recovered from the highest antibiotic concentrations after 72 h incubation were passaged five times on drug-free MHA. Antibiotic MICs against putative mutants were retested before and after passage on drug-free media. If any antibiotic MIC against mutants selected from 1/2 MPC was below the resistant breakpoint concentration, MHA plates containing antibiotic concentrations of 50%$\times$, 60%$\times$, 70%$\times$, 80%$\times$, 90%$\times$ and 100%$\times$MPC were prepared and inoculated with 1.2×10^{10} cells of MRSE to see whether higher antibiotic MIC against mutants (doubling dilutions) would result. Retesting putative mutant for susceptibility should be accompanied by MIC retesting of its parental strain for comparison.

RESULTS

All MICs were conducted in triplicate on separate days. Table 1 lists MICs of the four antimicrobial agents mentioned above against 10 strains of MRSE. The susceptibility rates of these strains to vancomycin, rifampicin, levofloxacin and fosfomycin were 100%, 100%, 50% and 90%, respectively.

The FICI of vancomycin combination with rifampicin, levofloxacin or fosfomycin were $\geqslant 1.5$ but $\leqslant 2$ for 10 parental strains, $\geqslant 1.5$ but $\leqslant 2$ for 5 parental strains and >0.5 but $\leqslant 1.5$ for 9 parental strains, respectively.

All MPC determinations were made in duplicate and the results were identical. Table 2 reflects the MPCs of antimicrobials alone and in combination with vancomycin against MRSE isolates. Retesting putative mutants selected by individual antibiotics for susceptibility demonstrated that antibiotics had standard MICs of $>64\,\mu g\,ml^{-1}$ for rifampicin, 4–8 $\mu g\,ml^{-1}$ for levofloxacin and $\geqslant 64\,\mu g\,ml^{-1}$ for

Table 1 MICs of four antimicrobials against 10 strains of MRSE ($\mu g\,ml^{-1}$)

	MIC for			
Isolate No.	*VAN*	*RIF*	*LVX*	*FOF*
1	1	0.008	0.25	16
2	1	0.008	0.25	0.25
3	0.5	0.008	0.25	0.5
4	0.5	0.008	2	1
5	0.5	0.008	8	0.5
6	0.5	0.008	>32	1
7	0.5	0.015	8	64
8	1	0.008	0.125	2
9	0.5	0.008	0.5	1
10	0.5	0.015	>16	1

Abbreviations: FOF, fosfomycin; LVX, levofloxacin; MRSE, methicillin-resistant *Staphylococcus epidermidis;* RIF, rifampicin; VAN, vancomycin.

Table 2 MPCs of antimicrobials alone and in vancomycin combination against 10 strains of MRSE ($\mu g\,ml^{-1}$)

	MPC for						
Isolate No.	*VAN*	*FOF*	*VAN+FOF*	*LVX*	*VAN+LVX*	*RIF*	*VAN+RIF*
1	32	>64	16	8	32	>64	32
2	32	>64	1	4	16	>64	32
3	32	64	2	8	32	>64	32
4	16	64	2	NT[a]	NT[a]	>64	16
5	32	>64	16	NT[a]	NT[a]	>64	32
6	16	64	4	NT[a]	NT[a]	>64	16
7	16	NT[a]	NT[a]	NT[a]	NT[a]	>64	16
8	32	>64	2	4	32	>64	32
9	32	64	4	16	32	>64	32
10	32	>64	4	NT[a]	NT[a]	>64	16

Abbreviations: FOF, fosfomycin; LVX, levofloxacin; MPC, mutant-prevention concentration; MRSE, methicillin-resistant *Staphylococcus epidermidis;* RIF, rifampicin; VAN, vancomycin.
[a]NT, not tested.

fosfomycin. In contrast, against mutants selected by vancomycin individually, vancomycin had MICs of $2\,\mu g\,ml^{-1}$ (for parental isolates 2–4 and 6–10) or $4\,\mu g\,ml^{-1}$ (for parental isolates 1 and 5). Higher vancomycin concentrations than 1/2 MPC (that is, 60%×, 70%×, 80%× or 90%×MPC) did not result in higher vancomycin MIC against selected mutants (doubling dilutions). All the antibiotic MIC values against selected mutants remain unchanged before and after 5 consecutive days of passaging on drug-free agar plates. Antibiotic MICs against the parental strains remained the same as previously determined.

DISCUSSION

The inoculum effect here refers to antibiotic efficacy being impaired by large inocula (density), such as density-dependent declines in the rate and extent of antibiotic-mediated killing and increases in MIC.[25] There have been other *in vitro* and animal studies that support the inoculum effect.[26–28] Likewise, MPC increases with bacterial density because of the inoculum effect.[29] Thus, antibiotic MPCs determined on agar plates are likely to be higher than the standard low-inoculum antibiotic MIC against the least susceptible single-step mutant—that is, MPC should be regarded as the actual MIC rather than the theoretically defined standard low-inoculum MIC against the least susceptible single-step mutants.

The findings indicate that single-step mutation cannot result in vancomycin non-susceptibility in MRSE (MIC $>4\,\mu g\,ml^{-1}$), which is consistent with previous studies and the fact that non-susceptibility of staphylococci to vancomycin arises rarely in the clinic.[12,13,30] In contrast, single-step resistant mutants could be selected relatively easily by exposure to antibacterials other than vancomycin, which is consistent with the fact that resistance of staphylococci to rifampicin, levofloxacin or fosfomycin is relatively common in the clinic. However, it should be realized that two factors account for the fact that resistance selection *in vivo* is usually more difficult than that *in vitro*: one is that the antibiotic concentration *in vivo* is not high enough to select resistance, the other is that the immune system can kill mutants effectively. According to this study, infections with a large enough bacterial density (for example, deep-seated abscess and native valve endocarditis) might not be eradicated because of the inoculum effect unless vancomycin concentration at infection sites reaches the elevated MPCs, but such a target can hardly be accomplished at infection sites other than serum. Besides, mutants can be selected and enriched where vancomycin penetrates poorly (for example, the lung, cerebrospinal fluid and bone joint) and when vancomycin dosing is suboptimal (for example, for patients undergoing dialysis therapy). All these conditions make it necessary to apply combination therapy instead of monotherapy.

Fortunately, compared with vancomycin alone, combined application of vancomycin and fosfomycin ($32\,\mu g\,ml^{-1}$) resulted in seven MPC reductions by at least four-fold within the susceptibility range of *S. epidermidis* to vancomycin ($\leqslant 4\,\mu g\,ml^{-1}$), with the exceptions of isolates 1 and 5. Although vancomycin combination with fosfomycin showed indifferent interaction in synergy testing in this study and the previous study,[31] considering mutant growth can be blocked by concurrent use of vancomycin and fosfomycin at concentrations below their respective MPCs, an additive protective effect of vancomycin plus fosfomycin combinations on the selection of resistant *S. epidermidis* may be proposed. This slows the emergence of vancomycin non-susceptibility because the selective enrichment of vancomycin-susceptible mutants (that is, precursors of vancomycin non-susceptible mutants) is inhibited. An obvious fact was that against the least susceptible single-step mutants selected by vancomycin from parental isolates 1 and 5, vancomycin had MICs identical to the susceptible breakpoint concentration of *S. epidermidis* to vancomycin. This might be the reason why the vancomycin MPCs against these two isolates were not reduced by more than two-fold in combination with fosfomycin. Surprisingly, vancomycin in combination with rifampicin (2 or $4\,\mu g\,ml^{-1}$) or levofloxacin ($2\,\mu g\,ml^{-1}$) did not lead to significant lowering of the isolates' vancomycin MPCs in general. According to Zhanel *et al.*[32] and Firsov *et al.*,[33] the MPC of an antibiotic could not be lowered when combined with a partner at a relatively lower concentration. Likewise, the failure of vancomycin MPCs to be significantly lowered when combined with levofloxacin or rifampicin might also result from the lower concentrations of levofloxacin and rifampicin used in combination (levofloxacin and rifampicin are both concentration-dependent antibiotics). The reason for such a phenomenon was unanimously explained as an emergence of cross resistance by Zhanel *et al.*[32] and Firsov *et al.*[33] Here, we give a possible explanation from another perspective. Theoretically speaking, two-step mutations would be required for mutant growth once both drugs are kept above MIC. However, because of the inoculum effect, the actual MIC

(not the standard low-inoculum MIC) for an antibiotic increases with bacterial density. Thus, single-step mutation would result in mutant growth if one drug is kept above its actual MIC, whereas the other is below its actual MIC.

A recent study using a biofilm model of methicillin-resistant *S. aureus* by Tang *et al.*[34] has confirmed that the antibacterial activity of vancomycin combination with fosfomycin is superior to vancomycin individually and vancomycin in combination with rifampin. Similar conclusion might also be applicable for MRSE. Forrest and Tamura[9] retrospectively analyzed previous studies *in vitro* and *in vivo*. They concluded that rifampicin combination therapy seems to have improved treatment outcomes when there is a low-organism-burden infection such as those with biofilms (that is, prosthetic joint infections and prosthetic valve endocarditis caused by methicillin-susceptible staphylococci) but in general does not offer any benefit over antibiotic monotherapy for high-organism-burden infections such as native valve endocarditis. Time-kill studies by Kang-Birken[35] confirmed the bactericidal activity of vancomycin combination with levofloxacin mimicked that of vancomycin alone. All these studies do not contradict our study.

CONCLUSION

In all, monotherapy of rifampicin, levofloxacin or fosfomycin is not recommended for MRSE infections because they cannot block the proliferation of resistant mutants. Likewise, suboptimal dosing of vancomycin monotherapy can also result in enrichment of mutants and decreased susceptibility of MRSE to vancomycin. Concurrent use of fosfomycin and vancomycin is probably a superior alternative when and where vancomycin monotherapy is not sufficient, especially for high-organism-burden infections. Furthermore, available data indicate that the combinational application of vancomycin and fosfomycin alleviates the renal toxicity caused by vancomycin.[24] In contrast, vancomycin combination with rifampicin or levofloxacin is not recommended for MRSE infections. Our conclusions need to be validated *in vivo*.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (no. 81071394).

1 Bereket, W. *et al.* Update on bacterial nosocomial infections. *Eur. Rev. Med. Pharmacol. Sci.* **16,** 1039–1044 (2012).
2 Piette, A. & Verschraegen, G. Role of coagulase-negative staphylococci in human disease. *Vet. Microbiol* **134,** 45–54 (2009).
3 Karl, D. The mutant selection window and antimicrobial resistance. *J. Antimicrob. Chemother.* **52,** 11–17 (2003).
4 Zhao, X. & Drlica, K. Restricting the selection of antibiotic-resistant mutants: a general strategy derived from fluoroquinolone studies. *Clin. Infect. Dis.* **33** (Suppl 3), S147–S156 (2001).
5 Srinivasan, A., Dick, J. D. & Perl, T. M. Vancomycin resistance in staphylococci. *Clin. Microbiol. Rev.* **15,** 430–438 (2002).
6 Szabó, J. hVISA/VISA: diagnostic and therapeutic problems. *Expert Rev. Anti. Infect. Ther.* **7,** 1–3 (2009).
7 Shoji, M. *et al.* walK and clpP mutations confer reduced vancomycin susceptibility in *Staphylococcus aureus. Antimicrob. Agents Chemother.* **55,** 3870–3881 (2011).
8 Smith, H. J., Nichol, K. A., Hoban, D. J. & Zhanel, G. G. Stretching the mutant prevention concentration (MPC) beyond its limits. *J. Antimicrob. Chemother.* **51,** 1323–1325 (2003).
9 Forrest, G. N. & Tamura, K. Rifampin combination therapy for nonmycobacterial infections. *Clin. Microbiol. Rev.* **23,** 14–34 (2010).
10 Michalopoulos, A. S., Livaditis, I. G. & Gougoutas, V. The revival of fosfomycin. *Int. J. Infect. Dis.* **15,** e732–e739 (2011).
11 Falagas, M. E. *et al.* Antimicrobial susceptibility of Gram-positive non-urinary isolates to fosfomycin. *Int. J. Antimicrob. Agents* **35,** 497–499 (2010).
12 Zhao, C. *et al.* Antimicrobial resistance trends among 5608 clinical Gram-positive isolates in China: results from the Gram-Positive Cocci Resistance Surveillance program (2005-2010). *Diagn. Microbiol. Infect. Dis.* **73,** 174–181 (2012).
13 Hope, R., Livermore, D. M., Brick, G., Lillie, M. & Reynolds, R BSAC Working Parties on Resistance. Non-susceptibility trends among staphylococci from bacteraemias in the UK and Ireland, 2001-06. *J. Antimicrob. Chemother.* **62** (Suppl 2), ii65–ii74 (2008).
14 Zhang, K. *et al.* New quadriplex PCR assay for detection of methicillin and mupirocin resistance and simultaneous discrimination of *Staphylococcus aureus* from coagulase-negative staphylococci. *J. Clin. Microbiol.* **42,** 4947–4955 (2004).
15 Martineau, F., Picard, F. J., Roy, P. H., Ouellette, M. & Bergeron, M. G. Species-specific and ubiquitous-DNA-based assays for rapid identification of *Staphylococcus epidermidis. J. Clin. Microbiol.* **34,** 2888–2893 (1996).
16 Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing; twenty-second informational supplement. *CLSI document M100-S22* **32,** 1–188 (2012).
17 European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. *Version 2.0* (2012). http://www.eucast.org/clinical _breakpoints (5 January 2012, date last accessed).
18 Lozniewski, A., Lion, C., Mory, F. & Weber, M. *In vitro* synergy between cefepime and vancomycin against methicillin-susceptible and -resistant *Staphylococcus aureus* and *Staphylococcus epidermidis. J. Antimicrob. Chemother* **47,** 83–86 (2001).
19 Odds, F. C. Synergy, antagonism, and what the chequerboard puts between them. *J. Antimicrob. Chemother.* **52,** 1 (2003).
20 Zhao, X., Eisner, W., Perl-Rosenthal, N., Kreiswirth, B. & Drlica, K. Mutant prevention concentration of garenoxacin (BMS-284756) for ciprofloxacin-susceptible or -resistant *Staphylococcus aureus. Antimirob. Agents Chemother.* **47,** 1023–1027 (2003).
21 Chien, S. C. *et al.* Pharmocokinetic profile of levofloxacin following once-daily 500-milligram oral or intravenous doses. *Antimirob. Agents Chemother.* **41,** 2256–2260 (1997).
22 Ruslami, R. *et al.* Pharmacokinetics and tolerability of a higher rifampin dose versus the standard dose in pulmonary tuberculosis patients. *Antimirob. Agents Chemother.* **51,** 2546–2551 (2007).
23 Ruslami, R. *et al.* Pharmacokinetics of antituberculosis drugs in pulmonary tuberculosis patients with type 2 diabetes. *Antimirob. Agents Chemother.* **54,** 1068–1074 (2010).
24 Roussos, N., Karageorgopoulos, D. E., Samonis, G. & Falagas, M. E. Clinical significance of the pharmacokinetic and pharmacodynamic characteristics of fosfomycin for the treatment of patients with systemic infections. *Int. J. Antimicrob. Agents* **34,** 506–515 (2009).
25 Udekwu, K. I., Parrish, N., Ankomah, P., Baquero, F. & Levin, B. R. Functional relationship between bacterial cell density and the efficacy of antibiotics. *J. Antimicrob. Chemother.* **63,** 745–757 (2009).
26 Bhagunde, P. *et al.* Mathematical modeling to characterize the inoculum effect. *Antimicrob. Agents Chemother.* **54,** 4739–4743 (2010).
27 Soriano, F., Santamaría, M., Ponte, C., Castilla, C. & Fernández-Roblas, R. *In vivo* significance of the inoculum effect of antibiotics on *Escherichia coli. Eur. J. Clin. Microbiol. Infect. Dis.* **7,** 410–412 (1988).
28 Soriano, F., Ponte, C., Santamaría, M. & Jimenez-Arriero, M. Relevance of the inoculum effect of antibiotics in the outcome of experimental infections caused by *Escherichia coli. J. Antimicrob. Chemother.* **25,** 621–627 (1990).
29 Quinn, B., Hussain, S., Malik, M., Drlica, K. & Zhao, X. Daptomycin inoculum effects and mutant prevention concentration with *Staphylococcus aureus. J. Antimicrob. Chemother.* **60,** 1380–1383 (2007).
30 Biavasco, F., Vignaroli, C. & Varaldo, P. E. Glycopeptide resistance in coagulase-negative staphylococci. *Eur. J. Clin. Microbiol. Infect. Dis.* **19,** 403–417 (2000).
31 Grif, K., Dierich, M. P., Pfaller, K., Miglioli, P. A. & Allerberger, F. *In vitro* activity of fosfomycin in combination with various antistaphylococcal substances. *J. Antimicrob. Chemother.* **48,** 209–217 (2001).
32 Zhanel, G. G., Mayer, M., Laing, N. & Adam, H. J. Mutant prevention concentrations of levofloxacin alone and in combination with azithromycin, ceftazidime, colistin (Polymyxin E), meropenem, piperacillin-tazobactam, and tobramycin against *Pseudomonas aeruginosa. Antimicrob. Agents Chemother.* **50,** 2228–2230 (2006).
33 Firsov, A. A., Vostrov, S. N., Lubenko, I. Y., Portnoy, Y. A. & Zinner, S. H. Prevention of the selection of resistant *Staphylococcus aureus* by moxifloxacin plus doxycycline in an *in vitro* dynamic model: an additive effect of the combination. *Int. J. Antimicrob. Agents* **23,** 451–456 (2004).
34 Tang, H. J. *et al. In vitro* efficacy of fosfomycin-containing regimens against methicillin-resistant *Staphylococcus aureus* in biofilms. *J. Antimicrob. Chemother.* **67,** 944–950 (2012).
35 Kang-Birken, S. L. Comparative *in vitro* activity of vancomycin and levofloxacin in combination with rifampin against planktonic versus sessile cells of *Staphylococcus epidermidis. Pharmacotherapy* **20,** 673–678 (2000).

PROGRESS

Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance

Adam C. Palmer and Roy Kishony

Abstract | The evolution of antibiotic resistance can now be rapidly tracked with high-throughput technologies for bacterial genotyping and phenotyping. Combined with new approaches to evolve resistance in the laboratory and to characterize clinically evolved resistant pathogens, these methods are revealing the molecular basis and rate of evolution of antibiotic resistance under treatment regimens of single drugs or drug combinations. In this Progress article, we review these new tools for studying the evolution of antibiotic resistance and discuss how the genomic and evolutionary insights they provide could transform the diagnosis, treatment and predictability of antibiotic resistance in bacterial infections.

The evolution and spread of antibiotic resistance in bacterial pathogens is a growing threat to public health. The frequency of antibiotic resistance in many bacterial pathogens is increasing around the world, and the resulting failures of antibiotic therapy cause hundreds of thousands of deaths annually[1]. The hope of addressing this crisis by developing new antibiotics is diminished both by the low rate of novel antibiotic discovery and by the likelihood that pathogens will evolve resistance to novel antibiotics just as they have to existing antibiotics. The long-term threat, therefore, is just as much the process of evolution as the microbial pathogens themselves. Although the use of antibiotics inevitably promotes resistance, the rate of evolution depends on the genomic background and treatment strategies. Thus, understanding the genomics and evolutionary biology of antibiotic resistance could inform therapeutic strategies that are both effective and mitigate the future potential to evolve resistance.

Antibiotic resistance can be acquired either by mutation or by the horizontal transfer of resistance-conferring genes, often in mobile genetic cassettes. The relative contribution of these factors depends on the class of antibiotic and on the genetic plasticity of the bacterial species. For example, *Mycobacterium tuberculosis* primarily acquires antibiotic resistance through nucleotide changes, whereas hospital-acquired Enterobacteriaceae infections often possess multidrug resistance cassettes and may also acquire nucleotide changes that confer resistance to drugs that are not often resisted by mobile elements, such as quinolones[2].

Progress in DNA sequencing and other genotyping technologies means that the genotypes of pathogens will soon be widely available in clinical as well as research settings. Genotype-based antibiotic resistance profiling is already faster and more economical than phenotypic profiling in select cases (for example, rifampicin resistance in *M. tuberculosis* caused by nucleotide substitutions, and methicillin resistance in *Staphylococcus aureus* caused by a resistance cassette), and over time therapeutic and infection control strategies will more heavily rely on information derived from genome sequencing of the infecting agents[3].

Importantly, genotypes can inform not only on the current drug susceptibility of a pathogen but also on its future potential to evolve resistance and spread. For example, sequencing could determine whether a drug-susceptible strain carries precursors to resistance genes (which are termed proto-resistance genes[4]), such as drug-degrading enzymes or efflux pumps, that might be mutated to increase expression or to strengthen activity. Sequencing could also determine whether resistance cassettes may be only one mutation away from increased potency or from the capacity to resist other drugs related to the originally resisted drug[4,5]. Even a modest predictive power might improve therapeutic outcomes by informing the selection of drugs, the preference between monotherapy or combination therapy and the temporal dosing regimen to select genotype-based treatments that are most resilient to evolution of resistance. To realize such a potential will require new tools to explore how different treatment regimes affect the genotypic and phenotypic evolutionary paths to antibiotic resistance in the laboratory and in the clinic. Here we discuss new tools to select for drug resistance, strategies for identifying and characterizing adaptive mutations in the evolved genotypes, and approaches to study the genetic constraints on the evolution of resistance.

Selection for drug resistance

Drug resistance in laboratory experiments. Laboratory evolution[6] can investigate how the rate and genotypic path to resistance varies across different controlled drug treatment regimens. In a traditional selection experiment, bacteria are exposed to fixed drug doses that permit only the growth of resistant mutants. Typically, this approach identifies only a single adaptive step and does not reveal how multiple mutations can accrue sequentially to confer strong resistance (FIG. 1a). Technological innovations now facilitate rapid multistep experimental evolution, revealing long-term evolutionary paths. Recurrent evolutionary patterns, such as the appearance of mutations in a preferred order, provide some level of predictability to a seemingly stochastic evolutionary process[7]. Devices for establishing spatial or temporal gradients of drug concentration allow evolving populations to be continuously challenged by effectively increasing the drug dosage to maintain selective pressure

as stronger antibiotic resistance evolves. Continuous culture devices (for example, turbidostats) can be modified to increase drug dose steadily over time[8], to implement automated feedback control of drug dosage in response to increasing levels of resistance[7] or to mimic the antibiotic dosing regime experienced within a patient (FIG. 1b). Multistep experimental evolution can also be carried out in spatial drug gradients, as was demonstrated by a microfluidic device of connected chambers implementing a spatial drug gradient, allowing bacteria to expand throughout the device only as they evolve increasing levels of antibiotic resistance[9] (FIG. 1c). These experiments have revealed that although the evolution of resistance can follow similar phenotypic paths in replicate experiments, the underlying genotypic process can be variable for some drugs (for example, chloramphenicol and doxycycline) but reproducible for other drugs (for example, trimethoprim and ciprofloxacin)[7,9]. Substantial variability in rate is also observed: resistance to some drugs increases 1,000-fold over 20 days, whereas resistance to other drugs might increase only tenfold over the same period[7]. Therefore, for any specific genotype there could be vast differences between drugs in the propensity for resistance and the mechanisms by which resistance is acquired; these factors are crucial to the design of combination treatments that inhibit the evolution of resistance.

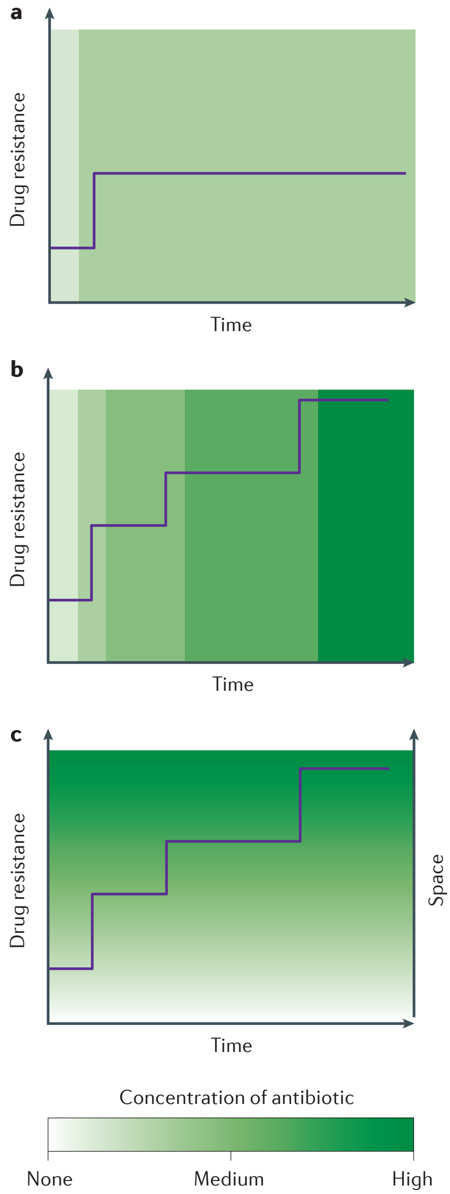


Figure 1 | **Selection of antibiotic-resistant bacteria from experimental evolution.** Gradients of drug concentration over time or space facilitate multistep experimental evolution. **a** | In a classical selection for antibiotic resistance, a uniform drug concentration selects for only a single mutation. **b** | A continuous culture device can select for multiple resistance-conferring mutations by dynamically increasing drug concentration in response to increasing drug resistance. **c** | If bacteria can migrate over a spatial gradient of drug concentration then they can explore larger regions of space only as they evolve increasing levels of drug resistance.

Combination therapy has the potential to slow the evolution of resistance, as a bacterial subpopulation with a mutation that renders it resistant to one drug may still be inhibited by a second drug, preventing the growth of a large drug-resistant population (that might subsequently evolve multidrug resistance)[10]. However, the choice of an optimal combination to slow evolution can crucially depend on the details of the treatment regimen, drug interactions and cross-resistance[10–13]. Experimental evolution has facilitated the systematic analysis of evolution under different combination therapies and is revealing the principles behind their ability to slow down and possibly even to reverse the evolution of resistance[12–15] (reviewed in REF. 16). Several approaches have been used to select for drug resistance in multidrug environments: mutants can be selected from a grid of drug concentrations across multiple agar dishes[12] or in a microtitre plate[13]. Multiple mutations that confer strong multidrug resistance can be selected by serial passaging across such gradients[13] or through the use of drug combinations in the continuous culture devices described above[7].

Many questions about the evolution of multidrug resistance remain, including: to what extent is resistance acquired by a series of drug-specific mutations versus mutations that each confer resistance to multiple drugs (that is, positive cross-resistance); in which cases can resistance to one drug lead to sensitivity to another (that is, negative cross-resistance); and, even when resistance to one drug does not immediately confer positive or negative cross-resistance to a second drug, can it affect the future evolution of resistance to the second drug? These and other questions about the evolution of resistance to single or multiple drug treatments are being addressed by the systematic selection methodologies outlined above. Although these methods are often first applied to model organisms, they can and should be more widely applied to study pathogens isolated from human infections.

Drug resistance in clinical isolates. The increasing capacity to sequence whole bacterial genomes has allowed detailed analyses of large collections of clinical isolates. Various sampling approaches are available to view the evolution and spread of antibiotic-resistant bacteria over different scales (FIG. 2). Isolates collected from individual patients over the course of acute and chronic infections have revealed the within-patient evolution of antibiotic resistance, instances of cross-resistance between antibiotics, the evolution of compensatory mutations that alleviate the fitness costs of resistance, and the transmission of specific antibiotic-resistant clones between organs[17–19]. Sampling during the spread of an epidemic has been used to identify the likely patient-to-patient transmission of antibiotic-sensitive or antibiotic-resistant bacteria and may reveal trade-offs between infectivity and antibiotic resistance[18]. At the largest scale, worldwide sampling of endemic infections over decades has been used to determine long-term trends in the evolution of antibiotic resistance and pathogenicity and to determine transmission patterns across continents[20,21].

Finding the genotypic basis

Identifying adaptive mutations. Comparing the genomes of ancestral and evolved strains identifies the precise genetic changes that underlie adaptive evolution. However, separating adaptive mutations from neutral or passenger mutations is challenging, particularly for clinical strains that may have been evolving antibiotic resistance over decades. In the context of contemporary bacterial evolution, tests for adaptive evolution based on rates of nonsynonymous and synonymous substitutions (dN/dS) cannot be applied on a per-gene basis as there are typically too few mutations for statistical power; also, they should not be applied to a whole

genome because different subsets of genes may have undergone adaptive, neutral or purifying selection. Furthermore, such tests cannot determine whether an individual mutation is adaptive or neutral, and they neglect the possible role of adaptive non-coding or regulatory mutations.

Fortunately, with increasing capacity to sequence many evolved strains, this challenge can be overcome by looking for parallel evolution. Parallel evolution provides a tool to distinguish adaptive mutations from neutral or deleterious mutations, as non-advantageous mutations should not independently arise and fix at the same loci as frequently as adaptive mutations. Additionally, the identification of adaptive mutations by parallel evolution is not biased against synonymous or regulatory mutations, even though the set of adaptive mutations will probably be enriched for nonsynonymous substitutions. In a recent study of a bacterial epidemic in which parallel evolution occurred within multiple patients, the bacterial genome as a whole showed no statistical sign of adaptive evolution (the ratio of nonsynonymous to synonymous substitutions, dN/dS, was as expected under drift), but examining dN/dS identified adaptive evolution against a background signal of purifying selection when genes were classified by whether they mutated only once or whether repeatedly across the cohort. Genes that mutated only once showed signs of purifying selection (that is, unexpectedly few nonsynonymous substitutions), and genes that repeatedly mutated showed a strong signature of adaptive evolution (that is, an unexpectedly high rate of nonsynonymous substitutions)[18]. An important caveat applies to the identification of parallel evolution in clinical isolates: the repeated observation of a mutation could be a result of shared ancestry and does not necessarily imply that the same mutation arose repeatedly; a phylogenetic tree must be constructed to estimate the number of independent mutational events at each locus in the strains' histories (FIG. 3).

Phylogenetic trees describe the evolutionary history of related strains, providing crucial contributions to understanding mutation, selection and transmission in the evolution of antibiotic resistance (see REF. 22 for a tutorial). A phylogenetic tree provides a view of transmission events across as large or as fine a scale as is represented by clinical isolates, from intercontinental transmission to transfer between the organs of a single patient. Specific genetic changes throughout the evolutionary history of bacterial strains can be correlated with the appearance of novel phenotypes, including antibiotic resistance, changes in pathogenicity or fitness or propensity for transmission. Whereas recombination among related bacterial strains can complicate the construction of a phylogenetic tree, maximum likelihood and Bayesian approaches can identify clusters of mutations that are more likely to be shared by recombination than point mutation[23]. Phylogenetic reconstruction can then be carried out only on vertically transmitted point mutations, as demonstrated in a recent study of worldwide isolates of the highly recombinogenic *Streptococcus pneumoniae*[21].

Measuring the phenotypic effects of mutations. Even when adaptive mutations are identified, it is not necessarily straightforward to determine their specific phenotypic effects: that is, whether they increase antibiotic resistance, compensate for the fitness costs of antibiotic resistance or confer adaptation to a host environment. Thus, the lessons of high-throughput genotyping are limited unless combined with high-throughput phenotyping.

Fitness costs are a common feature of mutations that confer antibiotic resistance, which epidemiological models predict to substantially affect the spread of drug-resistant pathogens[24,25]. The relative fitness of evolved versus ancestral strains can be measured by competition experiments in drug-free or antibiotic-containing environments. Throughput and precision were previously limited by the labour of counting colonies, but these experiments can now be automated using fluorescent labels for counting by flow cytometry or DNA barcodes for counting by next-generation sequencing[14,15,26]. Similar methods can also be applied to genetically intractable clinical isolates by deep-sequencing of mutated loci to measure allele frequencies[27]. Improvements in the precision of fitness measurements will probably be of benefit to epidemiological modelling[25]. Another high-throughput phenotyping tool, which is applicable to model organisms and clinical isolates alike, is automated imaging arrays built from flatbed scanners. These cheap custom systems acquire time-lapse videos of colony growth on large numbers of agar plates that can be arranged to span ranges of antibiotic concentration or multiple antibiotics[12,28]. Studies using genetic complementation have also benefited from technological progress: the relative contributions to fitness of each mutated locus in an evolved strain can be simultaneously determined by a competition experiment between a mixture of strains, each transduced with a different fragment of the evolved genome[29]. Repeating such an experiment in the presence and absence of antibiotic could reveal both the degree of antibiotic resistance conferred by each mutation and the fitness cost during drug-free growth.

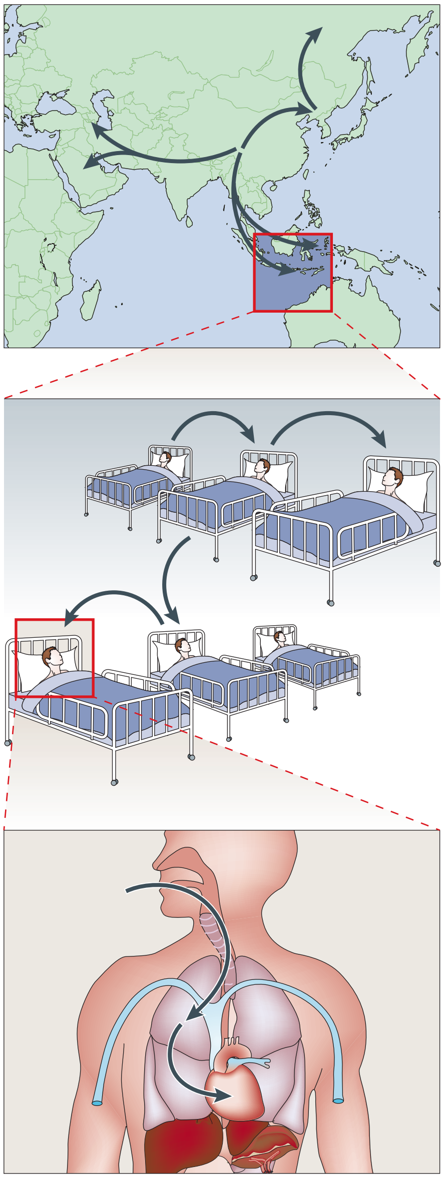


Figure 2 | **Selection of antibiotic-resistant bacteria from clinical isolates.** The evolution and transmission of antibiotic-resistant bacteria can be studied over scales ranging from continents to organs by different approaches from clinical sampling. Worldwide sampling of isolates reveals intercontinental transmission, sampling within a localized epidemic reveals patient-to-patient transmission networks, and sampling within a single patient can reveal transfer between sites of the body and possibly organ-specific evolution.

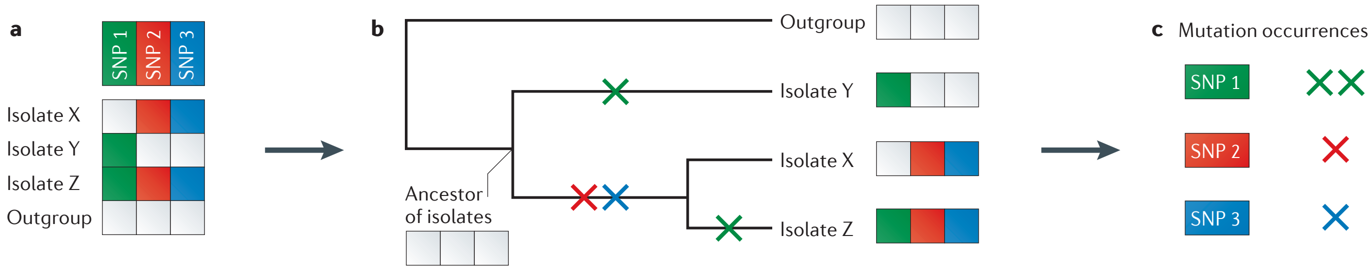


Figure 3 | **Phylogenetic inference identifies parallel evolution.** **a** | A collection of related isolates will possess many shared mutations relative to a more distant strain (an outgroup), but this does not necessarily imply that any of these mutations repeatedly occurred. **b** | Phylogenetic inference estimates the likely evolutionary history that connects the isolates and identifies when each mutation occurred. Note that many other mutations would need to have occurred for accurate phylogenetic inference; in this example, only three mutations are shown to illustrate the principle. **c** | From the phylogenetic tree, the number of times that a gene independently mutated in separate lineages can be counted to distinguish mutations that are shared merely by common ancestry (red and blue) from mutations that are shared by parallel evolution (green), strongly indicating adaptive evolution. SNP, single-nucleotide polymorphism.

Evolutionary potential and constraints

The approaches described above are based on natural selection methodologies to identify adaptive mutations that spontaneously appear under drug treatments. Such evolution-based approaches are powerful for determining the rate of adaptation and for revealing its most likely genotypic paths, but they do not explicitly elucidate unlikely or 'forbidden' steps that can have the effect of directing evolution repeatedly along the few permitted paths. To systematically explore the effects of defined genetic changes or combinations thereof, whether advantageous or deleterious, a reverse-genetics approach can be used. Here we review recent creative uses of reverse genetics to explore how systematic genetic perturbations, mutation combinations and horizontal gene transfer can enhance or constrain evolutionary potential.

Systematic genetic perturbations. The genetic determinants of antibiotic resistance can be explored with pre-constructed libraries of mutant strains. For example, known resistance genes can be mutagenized to explore their adaptive potential and to measure the distribution of mutational effects. Applying this method to the most common β-lactamase gene in Gram-negative bacteria, TEM-1, has identified a long-tailed distribution with a few highly beneficial mutations[30], potentially explaining the high degree of reproducibility often observed in the evolution of antibiotic resistance[7,9,31]. A genome-wide view can be taken with gene deletion libraries and open reading frame expression libraries; although these were first constructed only for model organisms, advances in transposon mutagenesis are making possible the rapid construction of comparable libraries for clinically relevant pathogens[32]. These libraries can be screened in pools by using next-generation sequencing to count the abundance of each strain in a mixture following drug selection[32]. Screening mutant strain libraries under antibiotic treatment identifies genes for which deletion or overexpression alters drug susceptibility, revealing the genetic basis of intrinsic antibiotic susceptibility or identifying paths to stronger antibiotic resistance[33]. Future studies could also use these mutant strain libraries as starting material for pooled evolution experiments, thereby identifying not only the immediate effects of the genetic perturbations but also their effect on the potential to evolve yet higher levels of resistance.

Combinatorial genetic libraries. The effects of a mutation depend on the genetic background on which it arises. Genetic interactions between alleles impose constraints on the evolutionary pathways to antibiotic resistance, as a mutation may be beneficial only in the presence or absence of certain other mutations. The synthetic construction of different combinations of mutations that have previously been identified from the clinic or experimental evolution can reveal genetic constraints that would not be observed from studying only those mutation combinations favoured in nature. This method has been applied to genes found in resistance cassettes as well as drug target genes, both being cases in which resistance can be increased by repeated mutation of the same gene. These studies have consistently observed strong constraints that can be responsible for the repeatability, and hence predictability, of evolutionary pathways[34]. Genetic interactions have been observed to limit the possible pathways to a few select sequences of mutations[35,36] (FIG. 4) and to limit to the reversibility of evolution when switching between different drugs[37]. This approach has shown that, in certain combinations, resistance mutations can also act as compensatory mutations that alleviate one another's fitness costs, producing strongly drug-resistant or multidrug-resistant strains without substantial fitness costs[38–40]. Evolutionary experiments can also be carried out starting from different pre-built genotypes to investigate genetic influences on the reproducibility of evolution: one such study has revealed that different initial mutations in the TEM-1 β-lactamase gene can define the subsequent evolutionary pathways[31] (FIG. 4). These approaches could identify those genotypes with a greater or lesser potential to evolve resistance to particular drugs, which could be valuable in selecting genotype-specific treatments that avoid the most harmful evolutionary outcomes.

Horizontal transfer of environmental genes. The acquisition of resistance by horizontal gene transfer (HGT) provides evolutionary potential that cannot be predicted from the original (pre-transfer) genome of an organism. Instead, the potential for resistance by HGT can be investigated by sampling the extensive and ancient ability of genes in environmental or commensal microbes to resist a drug[41–43]. This approach has been implemented by extracting and cloning microbial DNA from soil samples or from human gut samples into a laboratory strain and plating on inhibitory concentrations of a range of antibiotics to identify novel microbial drug resistance genes that might in the future transfer into pathogens[42,44]. Although this approach is limited to identifying genes that can be successfully expressed in the laboratory strain, a more recent study demonstrated an expression-independent approach that directly assessed the capacity of environmental microbes to degrade the

new and rarely clinically resisted antibiotic daptomycin[45]. A collection of environmental actinomycetes was screened for daptomycin resistance, and the supernatants of resistant cultures were analysed by mass spectrometry to view the structures of daptomycin and its inactivation products. By precisely viewing the drug degradation products, the molecular mechanisms of resistance by degradation were inferred. This level of understanding has the potential to suggest structural variants of drugs that could resist environmental mechanisms of degradation.

Towards therapies informed by evolution

The short- to medium-term challenge is to develop novel antibiotics to kill today's antibiotic-resistant bacteria, but the 'arms race' between antibiotic development and the evolution of resistance in pathogens is growing increasingly difficult. Addressing the long-term challenge posed by antibiotic-resistant pathogens will require therapeutic strategies[10] and compound development[46,47] that consider ways to manipulate and to slow the evolution of resistance.

The capacity for genome sequencing is approaching the point at which endemic pathogens can be extensively sampled and sequenced around the world, and evolution during bacterial outbreaks can be tracked in real time[48–50]. The ability to trace routes of bacterial transmission precisely is clearly of benefit to infection control efforts, which are especially important for highly drug-resistant infections for which few therapeutic options remain. Yet the value of genome sequencing extends beyond epidemic control: the genome of an organism defines its current antibiotic resistance and, to a large extent, its potential to evolve further resistance. The application of the methods reviewed here to genotype and to phenotype drug-resistant pathogens could identify resistance (or proto-resistance) genes and the ways that they might mutate to increase antibiotic resistance. Whole-genome sequencing of pathogens thus has the potential to provide a catalogue of various pathways to resistance under different treatment regimens.

To treat a drug-resistant infection, an understanding of interactions between drugs and resistance mutations can guide the selection of second-line therapies or combination therapies. Such therapies should have the weakest possible cross-resistance — negative cross-resistance if possible — and they should be chosen such that the genetic interactions between drug-specific resistance mutations lead to an accumulation of fitness

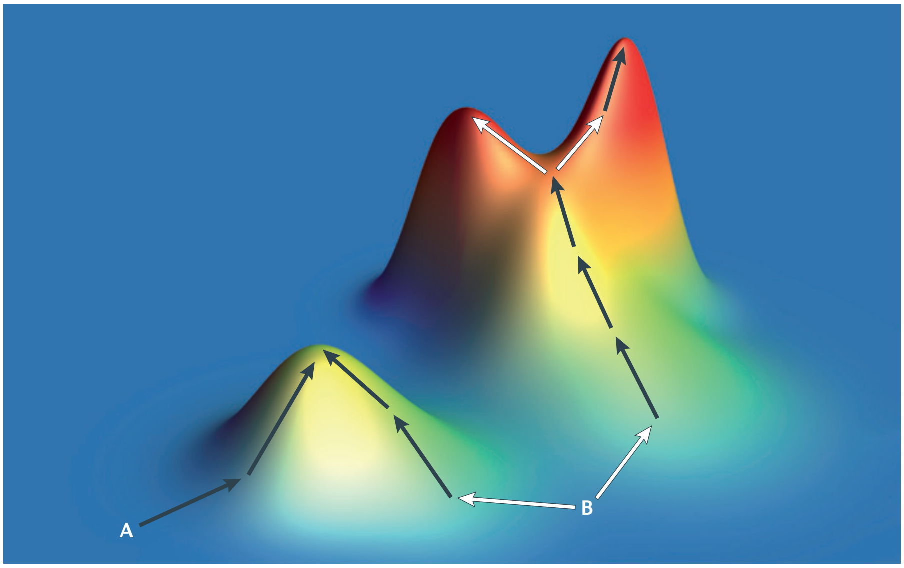


Figure 4 | **Constrained evolutionary pathways to antibiotic resistance.** The properties of evolutionary processes can be illustrated by the concept of the 'fitness landscape'. In this demonstration of several experimentally observed behaviours, height represents drug resistance. Different starting genotypes (A and B) may have a different propensity to evolve resistance owing to their proximities to drug-resistance peaks of varying height. The first genotypic step towards resistance can sometimes define the final genotype and the level of resistance (arrows from B). The pathways to resistance can at times be constrained and predictable (dark arrows), but evolutionary pathways can diverge (light arrows) to distinct peaks separated by negative genetic interactions.

Glossary

β-lactamase
An enzyme that can confer resistance to β-lactam antibiotics by catalysing their degradation.

Commensal microbes
Microbes living on or in a host without causing disease, although they typically include opportunistic pathogens.

Cross-resistance
The propensity of a genetic change that confers resistance to one drug also to affect resistance to a different drug (by either increasing or decreasing resistance).

dN/dS
The ratio of mutation rates at nonsynonymous (N) and synonymous (S) sites. dN/dS is increased by selection for amino acid changes (a signature of adaptive selection) and decreased by selection against amino acid changes (purifying selection).

Horizontal gene transfer
The acquisition of a gene by a means other than direct inheritance from a parent cell (vertical transfer). Common in many bacteria and archaea, mechanisms of horizontal gene transfer include transformation, conjugation and transduction.

Maximum likelihood and Bayesian approaches
This definition applies to the context of phylogenetics. Phylogenetic trees can be constructed by maximum parsimony, maximum likelihood and Bayesian inference. Maximum parsimony methods select from all possible trees the one containing the fewest mutations. Trees chosen by maximum likelihood and other Bayesian methods may contain more mutations, as they weigh the relative probabilities of different mutations according to various models.

Microfluidic device
Customized, microscopic chambers in which fluid flows can be precisely controlled. Applied to microbiology, these allow the study of bacterial behaviour in spatially and temporally controllable environments.

Monotherapy
Chemical therapy by a single drug.

Parallel evolution
When the same mutations (or a range of mutations in the same gene) repeatedly occur in independent lineages; this provides an indication that these mutations may have been fixed by positive selection rather than by chance.

Proto-resistance genes
Evolutionary precursors to drug-resistance genes that do not yet contribute to drug resistance but may do so on mutation and selection by drug stress.

Resistance cassettes
A genetic element containing one or more drug resistance genes, often carried in transposable elements or plasmids that facilitate horizontal gene transfer.

Transposon mutagenesis
The insertion of transposons at random locations throughout a genome to generate a library of different gene disruptions. Transposons can be constructed with outward-facing promoters also to introduce gene overexpression into the library.

Turbidostats
Devices that maintain constant cell density (turbidity) in a continuously growing microbial culture by routinely removing a small volume of culture and replacing it with fresh sterile media.

costs rather than compensation. Finally, the specific genetic basis of resistance to a first-line drug may, through genetic interactions, alter the expected capacity to evolve resistance to second-line drugs. Such cases might identify sets of drugs that, when used in sequence or in combinations, minimize the risk that strong resistance will evolve. The widespread application of the new methods reviewed in this article might thus facilitate an evolutionary medicine paradigm in which pathogen genotyping coupled with evolutionary genetics guides the optimal insights from choice of temporal and combinatorial drug treatment.

Adam C. Palmer is at the Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA.

Roy Kishony is at the Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA; the School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA; and the Faculty of Biology, Technion — Israel Institute of Technology, Haifa, Israel.

Correspondence to R.K.
e-mail: roy_kishony@hms.harvard.edu

doi:10.1038/nrg3351
Published online 19 February 2013

1. World Health Organization. *The evolving threat of antimicrobial resistance: options for action* (World Health Organization, 2012).
2. Paterson, D. L. Resistance in Gram-negative bacteria: Enterobacteriaceae. *Am. J. Med.* **119**, S20–S28; discussion S62–S70 (2006).
3. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. & Crook, D. W. Transforming clinical microbiology with bacterial genome sequencing. *Nature Rev. Genet.* **13**, 601–612 (2012).
4. Morar, M. & Wright, G. D. The genomic enzymology of antibiotic resistance. *Annu. Rev. Genet.* **44**, 25–51 (2010).
5. Novais, A. *et al.* Evolutionary trajectories of β-lactamase CTX-M-1 cluster enzymes: predicting antibiotic resistance. *PLoS Pathog.* **6**, e1000735 (2010).
6. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003).
7. Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature Genet.* **44**, 101–105 (2012).
8. Lee, H. H., Molla, M. N., Cantor, C. R. & Collins, J. J. Bacterial charity work leads to population-wide resistance. *Nature* **467**, 82–85 (2010).
9. Zhang, Q. *et al.* Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. *Science* **333**, 1764–1767 (2011).
10. Torella, J. P., Chait, R. & Kishony, R. Optimal drug synergy in antimicrobial treatments. *PLoS Comput. Biol.* **6**, e1000796 (2010).
11. Bonhoeffer, S., Lipsitch, M. & Levin, B. R. Evaluating treatment protocols to prevent antibiotic resistance. *Proc. Natl Acad. Sci. USA* **94**, 12106–12111 (1997).
12. Michel, J. B., Yeh, P. J., Chait, R., Moellering, R. C. Jr & Kishony, R. Drug interactions modulate the potential for evolution of resistance. *Proc. Natl Acad. Sci. USA* **105**, 14918–14923 (2008).
13. Hegreness, M., Shoresh, N., Damian, D., Hartl, D. & Kishony, R. Accelerated evolution of resistance in multidrug environments. *Proc. Natl Acad. Sci. USA* **105**, 13977–13981 (2008).
14. Chait, R., Craney, A. & Kishony, R. Antibiotic interactions that select against resistance. *Nature* **446**, 668–671 (2007).
15. Palmer, A. C., Angelino, E. & Kishony, R. Chemical decay of an antibiotic inverts selection for resistance. *Nature Chem. Biol.* **6**, 105–107 (2010).
16. Yeh, P. J., Hegreness, M. J., Aiden, A. P. & Kishony, R. Drug interactions and the evolution of antibiotic resistance. *Nature Rev. Microbiol.* **7**, 460–466 (2009).
17. Mwangi, M. M. *et al.* Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **104**, 9451–9456 (2007).
18. Lieberman, T. D. *et al.* Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genet.* **43**, 1275–1280 (2011).
19. Comas, I. *et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature Genet.* **44**, 106–110 (2012).
20. Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–474 (2010).
21. Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).
22. Baldauf, S. L. Phylogeny for the faint of heart: a tutorial. *Trends Genet.* **19**, 345–351 (2003).
23. Didelot, X. & Falush, D. Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**, 1251–1266 (2007).
24. Cohen, T. & Murray, M. Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness. *Nature Med.* **10**, 1117–1121 (2004).
25. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Rev. Microbiol.* **8**, 260–271 (2010).
26. Hegreness, M., Shoresh, N., Hartl, D. & Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**, 1615–1617 (2006).
27. Chubiz, L. M., Lee, M. C., Delaney, N. F. & Marx, C. J. FREQ-seq: a rapid, cost-effective, sequencing-based method to determine allele frequencies directly from mixed populations. *PLoS ONE* **7**, e47959 (2012).
28. Levin-Reisman, I. *et al.* Automated imaging with ScanLag reveals previously undetectable bacterial growth phenotypes. *Nature Methods* **7**, 737–739 (2010).
29. Goodarzi, H., Hottes, A. K. & Tavazoie, S. Global discovery of adaptive mutations. *Nature Methods* **6**, 581–583 (2009).
30. Schenk, M. F., Szendro, I. G., Krug, J. & de Visser, J. A. Quantifying the adaptive potential of an antibiotic resistance enzyme. *PLoS Genet.* **8**, e1002783 (2012).
31. Salverda, M. L. *et al.* Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **7**, e1001321 (2011).
32. van Opijnen, T., Bodi, K. L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods* **6**, 767–772 (2009).
33. Girgis, H. S., Hottes, A. K. & Tavazoie, S. Genetic architecture of intrinsic antibiotic susceptibility. *PLoS ONE* **4**, e5629 (2009).
34. Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).
35. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
36. Lozovsky, E. R. *et al.* Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl Acad. Sci. USA* **106**, 12025–12030 (2009).
37. Tan, L., Serene, S., Chao, H. X. & Gore, J. Hidden randomness between fitness landscapes limits reverse evolution. *Phys. Rev. Lett.* **106**, 198102 (2011).
38. Trindade, S. *et al.* Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet.* **5**, e1000578 (2009).
39. Brown, K. M. *et al.* Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. *Mol. Biol. Evol.* **27**, 2682–2690 (2010).
40. Hall, A. R. & MacLean, R. C. Epistasis buffers the fitness effects of rifampicin- resistance mutations in *Pseudomonas aeruginosa*. *Evolution* **65**, 2370–2379 (2011).
41. D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. Sampling the antibiotic resistome. *Science* **311**, 374–377 (2006).
42. Sommer, M. O., Church, G. M. & Dantas, G. The human microbiome harbors a diverse reservoir of antibiotic resistance genes. *Virulence* **1**, 299–303 (2010).
43. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
44. Riesenfeld, C. S., Goodman, R. M. & Handelsman, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* **6**, 981–989 (2004).
45. D'Costa, V. M. *et al.* Inactivation of the lipopeptide antibiotic daptomycin by hydrolytic mechanisms. *Antimicrob. Agents Chemother.* **56**, 757–764 (2012).
46. Chusri, S., Villanueva, I., Voravuthikunchai, S. P. & Davies, J. Enhancing antibiotic activity: a strategy to control *Acinetobacter* infections. *J. Antimicrob. Chemother.* **64**, 1203–1211 (2009).
47. Lewis, K. Antibiotics: recover the lost art of drug discovery. *Nature* **485**, 439–440 (2012).
48. Koser, C. U. *et al.* Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.* **366**, 2267–2275 (2012).
49. Snitkin, E. S. *et al.* Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci. Transl. Med.* **4**, 148ra116 (2012).
50. Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect. Dis.* **13**, 130–136 (2012).

Acknowledgements
We thank T. Lieberman for discussions on phylogeny and comments on the manuscript. This work was supported in part by US National Institutes of Health grants R01GM081617 and US National Institute of General Medical Science Center grant P50GM068763, and the Novartis Institutes for BioMedical Research.

Competing interests statement
The authors declare no competing financial interests.

COMMENT

ND4BB: addressing the antimicrobial resistance crisis

John H. Rex

The Innovative Medicines Initiative (IMI) recently launched its third public–private partnership, ENABLE (European Gram-negative Antibacterial Engine), to tackle the shortage of effective antimicrobial drugs for Gram-negative pathogens.

The extraordinary power of antibiotics has transformed modern medicine by saving countless lives, but every use of an antibiotic carries the risk of selecting for resistance, undermining its future use. Owing to the extensive use and misuse of the available antibiotic arsenal and the shortage of new drugs reaching the market, antimicrobial resistance (AMR) is now a major threat to healthcare worldwide. The alarm was raised 10 years ago by the Infectious Diseases Society of America in the first of their series of Bad Bugs, No Drugs papers[1]. Since then, the lack of an innovative and sustainable pipeline of new antimicrobial agents has been the subject of much discussion, and in recent years, several important initiatives to tackle the diminishing pipeline have been established.

ENABLE is the third project to be launched since the European Commission and its industry partners in the European Federation of Pharmaceutical Industries and Associations (EFPIA) began their ground-breaking New Drugs For Bad Bugs (ND4BB) campaign in May 2012. ND4BB was created by the IMI in response to the 'Action plan against the rising threats from Antimicrobial Resistance' and is the foundational infrastructure for a series of projects that have been initiated to establish an EU-wide plan to combat AMR. Each of the ND4BB projects is a public–private partnership in which pharmaceutical developers, academic researchers and small-to-medium-sized biotechnology companies collaborate under a framework that ensures data-sharing across the partners and across the wider research community.

Planned at the time to provide substantial funding to tackle the problem of AMR, ND4BB quickly launched the projects TRANSLOCATION and COMBACTE (Combating Bacterial Resistance in Europe). At the discovery level, TRANSLOCATION is an interdisciplinary initiative that examines how antibiotics are transported across the bacterial cell wall and membrane. This knowledge should improve our understanding of the molecular basis of resistance in Gram-negative pathogens and guide the design of small molecules that efficiently enter the cell via porins and avoid subsequent efflux. In addition, the project also explores the diversity of bacterial uptake pathways to further our understanding of how these pathways could be hijacked to deliver potent antibacterial agents into the cell. This research should provide important tools and knowledge that will be used to enhance the success of current and future drug discovery programmes. The work of COMBACTE is directed towards the development phase — in particular, the design and implementation of more efficient clinical trials through the creation of a clinical trial network (COMBACTE CLIN-Net), the creation of a European laboratory surveillance network (COMBACTE LAB-Net) and pioneering novel statistical methods and innovative clinical trial designs. COMBACTE also provides support for clinical trials of novel antimicrobial agents that have the potential to treat multidrug-resistant respiratory and skin pathogens, such as methicillin-resistant *Staphylococcus aureus* (MRSA).

The ENABLE project, which was announced by the IMI on 12 February 2014, is the latest addition to the ND4BB series. This 6-year project, which has secured €85 million in funding, involves 13 countries and 32 partners from academia, small–medium enterprises (SMEs) and the pharmaceutical industry. This project will create a drug discovery engine in which all of the stakeholders collaborate to fast-track the development of hits and leads into drug candidates to treat multidrug-resistant Gram-negative infections. By bringing together the experience and knowledge of academic scientists and drug discovery experts, the ultimate goal is to identify new agents that are active against resistant Gram-negative bacteria and to take at least one new compound through Phase I clinical studies. The project will have the capacity to maintain a portfolio of up to four early-stage and two later-stage preclinical discovery programmes and is currently evaluating the initial collection of project proposals. To ensure a robust pipeline moving forwards, new partners will be invited to join for the continued development of new programmes.

Vice President and Head of Infection, Global Medicines Development, AstraZeneca, 35 Gatehouse Drive, Waltham, Massachusetts 02451, USA.
e-mail: John.Rex@astrazeneca.com
doi:10.1038/nrmicro3245
Published online
10 March 2014

These timely public–private partnerships join other projects that are now underway both in Europe and in the United States. For example, prior to the creation of ND4BB, IMI initiated the ongoing RAPP-ID (Rapid Point-of-care test Platforms for Infectious Diseases) project, which aims to develop fast and reliable point-of-care tests for the detection of various pathogens by combining novel probes, sample preparation methods and ultra-sensitive detection methods. In the United States, the NIAID (National Institutes of Allergy and Infectious Diseases) is leading a comprehensive Antibacterial Resistance Program that offers an extensive range of preclinical and clinical services as well as units that are focused on Phase I studies, vaccine evaluation and more. NIAID recently created another clinical research network led by the ARLG (Antibacterial Resistance Leadership Group). As overseers of the network's scientific agenda, ARLG will develop, design, implement and manage clinical studies and trials that focus on AMR. Their research programme includes innovative trials of early-stage novel agents, examining ways to optimize the use of current agents, evaluating new diagnostic methods and optimizing best practices for the prevention of infection. In further work in the United States, the Biomedical Advanced Research and Development Authority (BARDA) has formed several key public–private partnerships to support the development of new antibiotics. Together, the support that BARDA and NIAID offer to developers makes it increasingly feasible for drug companies to invest in the discovery and development of new antimicrobial agents.

It is exciting to see the establishment of this broad range of projects, which together, provide a systematic approach to address two of the three core causes of the diminishing antibiotic pipeline; namely, that new antibiotics are hard to discover, they are hard to develop and their economics are unfavourable[2]. The need for global action to address this triad was reinforced by a major review surveying the entire challenge, from the epidemiology of AMR to the need for updated diagnostic tools[3]. First, the challenge of discovery is being addressed by collaborative projects such as TRANSLOCATION and ENABLE and by direct research support, such as that provided by the NIAID and BARDA. Second, the challenge of development is being addressed by research networks, such as those led by COMBACTE and ARLG. In parallel, extensive discussions have led to a broadening of the regulatory pathways that control the introduction of new drugs to market. In this respect, progress has been made to enable new agents to be developed based on smaller clinical efficacy data sets, thus permitting their development in advance of the epidemic spread of new forms of resistance, as well as to treat less common and highly resistant organisms[4]. Such progress has been facilitated by recognizing the power of the twin sciences of pharmacokinetics and pharmacodynamics, which, when fully utilized, accelerate the clinical testing phase: preclinical predications of efficacy based on these tools are sufficiently powerful that clinical programmes can be streamlined in those settings in which there is a high unmet medical need.

> “at least three more projects are in planning and the total committed budget of the combined projects now exceeds €600 million”

Finally, and with these efforts focused on discovery and development well underway, attention has more recently turned to addressing the unfavourable economics of antibiotics. Only a few steps have been taken so far on this front, the most notable of which is the extended marketing exclusivity provisions created by the GAIN (Generate Antibiotic Incentives Now) Act, passed in 2012 in the United States. For agents known as Qualified Infectious Diseases Products (QIDP), the GAIN Act provides an economic incentive to developers in the form of an additional 5 years of marketing exclusivity. ND4BB will soon begin to address the economics issue with a project that is currently entitled ‘Driving Re-investment in R&D and Responsible Use of Antibiotics’, which will focus on the development of new business models and economic strategies to reinvigorate the antibiotic pipeline. Improved diagnostic tools[5] will assist more effective antibiotic stewardship but are inadequate without new business models that address the tension between the global need to implement the use of newly developed drugs and the equally clear need for their appropriate use. The economic challenge is further exacerbated by the mismatch between the extraordinary value of antibiotics to the global community and the general expectation that these drugs will continue to be readily available at an inexpensive cost. In this context, ND4BB's indication that a project aimed at improving our current business models will be forthcoming is very exciting.

Will these steps be enough? Only the future will really tell, but the initiatives discussed here are certainly a positive step towards safeguarding the supply of effective drugs to meet our current and future needs. Importantly, these are not the last projects that the IMI plans to launch — at least three more projects are in planning and the total committed budget of the combined projects now exceeds €600 million. It is also clear that efforts aimed at preventing infections before they occur must continue, through the implementation of appropriate infection control procedures, using vaccines where available and, at the most fundamental level, ensuring access to clean food and water.

Nevertheless, we know that not all infections can be prevented and, as a result, modern medicine is not possible without effective antimicrobial agents. Although the crisis is far from resolved, the leadership of the European Commission are to be commended for their far-sighted approach to creating ND4BB and its projects, all of which provide hope that the global community will have access to an adequate pipeline of novel antimicrobial agents with which to address the challenge of AMR.

1. Boucher H. W. *et al.* *Clin. Infect. Dis.* **56**, 1685–1694 (2013).
2. Morel C. M. & Mossialos E. *BMJ* **340**, c2115 (2010).
3. Laxminarayan R. *et al.* *Lancet Infect. Dis.* **13**,1057–1098 (2013).
4. Rex J. H. *et al.* *Lancet Infect. Dis.* **13**, 269–275 (2013).
5. Caliendo A. M. *et al.* *Clin. Infect. Dis.* **57**, S139–S170 (2013).

Competing interests statement
The author declares competing interests: see Web version for details.

OPEN

Emerging Microbes and Infections (2014) 3, e3; doi:10.1038/emi.2014.3

www.nature.com/emi

REVIEW

Persisters, persistent infections and the Yin–Yang model

Ying Zhang[1,2]

Persisters are a small fraction of quiescent bacterial cells that survive lethal antibiotics or stresses but can regrow under appropriate conditions. Persisters underlie persistent and latent infections and post-treatment relapse, posing significant challenges for the treatment of many bacterial infections. The current definition of persisters has drawbacks, and a Yin–Yang model is proposed to describe the heterogeneous nature of persisters that have to be defined in highly specific conditions. Despite their discovery more than 70 years ago, the mechanisms of persisters are poorly understood. Recent studies have identified a number of genes and pathways that shed light on the mechanisms of persister formation or survival. These include toxin–antitoxin modules, stringent response, DNA repair or protection, phosphate metabolism, alternative energy production, efflux, anti-oxidative defense and macromolecule degradation. More sensitive single-cell techniques are required for a better understanding of persister mechanisms. Studies of bacterial persisters have parallels in other microbes (fungi, parasites, viruses) and cancer stem cells in terms of mechanisms and treatment approaches. New drugs and vaccines targeting persisters are critical for improved treatment of persistent infections and perhaps cancers. Novel treatment strategies for persisters and persistent infections are discussed.
Emerging Microbes and Infections (2014) **3,** e3; doi:10.1038/emi.2014.3; published online 8 January 2014

Keywords: mechanisms; persistence; persisters; treatment strategies

BACTERIAL PERSISTERS

The phenomenon of bacterial persisters was first discovered by Gladys Hobby[1] in 1942, when penicillin was found to kill 99% of a streptococcal culture, leaving 1% of the bacterial population intact. This surviving 1% of the bacterial population not killed by penicillin was subsequently termed 'persisters' by Joseph Bigger[2] in 1944. The original definition of persisters by Bigger refers to a small population of dormant or non-growing bacteria that have non-heritable tolerance to penicillin but have the capacity to regrow and remain susceptible to the same antibiotic. This definition has drawbacks that recently are becoming increasingly recognized for several reasons. First, earlier studies did not appreciate the heterogeneity of persisters,[2,3] and it is only recently that persisters are found to be quite heterogeneous.[4–10] Second, Bigger's definition of persisters does not specify antibiotic exposure time and the time required to resume growth upon removal of antibiotics and culture media involved in cultivation. In fact, persisters are found to be relative,[4] and the age of bacterial culture, the type of antibiotics, antibiotic concentrations, length of antibiotic exposure, medium composition and aeration during antibiotic exposure can all affect the level of persisters.[4,10–12] This means that persisters in one condition may not be persisters in another condition. Third, the current persister definition is based on growth in fresh medium,[2,13] often quantified via colony-forming unit assays in which the number of bacteria growing on agar plates or, less commonly, where growth in liquid medium is monitored. This persister definition has limitations as it excludes viable but non-culturable[14] bacteria or dormant bacteria, which do not readily grow under 'normal' culture conditions but can grow under some conditions (upon extended incubation in liquid medium[15] or changing medium composition[10] or addition of resuscitation factors[16]) and are clinically relevant as part of the persister continuum (see below). Thus, a new persister definition is required to address the above issues not covered by the current definition. The new definition of persisters can be as follows: persisters refer to genetically drug susceptible quiescent (non-growing or slow growing) organisms that survive exposure to a given cidal antibiotic or drug and have the capacity to revive (regrow or resuscitate and grow) under highly specific conditions (see above for conditions affecting persister counts). The definition of persisters may be extended or broadened to include cidal stresses in place of cidal antibiotics in which case antibiotics can be viewed as a type of the cidal stress.

A Yin–Yang model is proposed to describe a dynamic and complex heterogeneous bacterial population consisting of growing (Yang, in red) and non-growing persister cells (Yin, in black) that are in varying growth and metabolic states *in continuum*[5,8] and can interconvert *in vitro* and *in vivo* (Figure 1). This Yin–Yang model is compatible with the above new definition of persisters and can account for the heterogeneity of persisters. Although there may not be persisters in an actively growing log phase culture initially, when the growing population (Yang) reaches a certain age and density, a small population of non-growing or slowly growing persisters (Yin) can emerge and increase in numbers as the culture ages.[12] The persister population (Yin) is heterogeneous and composed of various subpopulations with varying metabolic states *in continuum* in varying hierarchy, from shallow to deep persisters, which can encompass viable but non-culturable and various dormant variants with or without morphological changes as part of the persister continuum. Persisters not killed by antibiotics could revert to replicating forms (reverters) or damaged persister forms, which under appropriate conditions may have varying degrees

[1]Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA and [2]Key Laboratory of Medical Molecular Virology of MOE/MOH, Department of Infectious Diseases, Huashan Hospital, Fudan University, Shanghai 200040, China
Correspondence: Y Zhang
E-mail: yzhang@jhsph.edu
Received 15 August 2013; revised 30 October 2013; accepted 26 November 2013

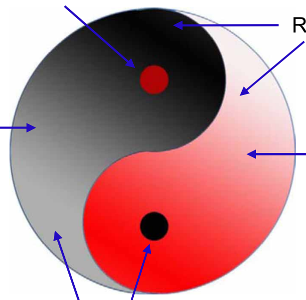


Figure 1 The Yin–Yang model of persisters and latent infections.[5,8,19] In a growing population of bacteria (Yang, red), there is a small population of non-growing or slowly growing persisters (Yin, black). In the persister population, there is a small number of growing bacteria (reverters). The persister population (Yin) or the growing population (Yang) is again heterogeneous and composed of various subpopulations with varying metabolic or dormant states *in continuum* in varying hierarchy (expressed by color from light to dark). The black spot in Yang (red) is connected to and the root of the Yin half (black), and the red spot in Yin, reverters, is connected to the Yang half (red). In the case of TB, INH kills growing bacteria (Yang) and RIF kills some growing bacteria, as well as slowly growing persisters, whereas PZA kills only persisters. Persisters not killed by antibiotics could revert to replicating forms (reverters) and cause relapse. The Yin–Yang model can be used to better describe latent infections (Yin) and active disease (Yang) at the host level and their respective interconversions.[8,19] As drug treatment and immune responses inhibit or kill the growing bacteria (Yang) and some of the persisters, some persisters (Yin) still remain and the infection becomes latent (Yin), but may revert and cause relapse or sustained chronic infections with symptoms. In a hierarchical manner and among heterogeneous persister cell populations, there are a few true 'stem' persister cells or mother cells (black spot in Yang) that have the capacity to derive other persisters (Yin) and initiate disease or cause reactivation. The Yin–Yang model proposes use of drugs targeting both replicating and non-replicating cells in combination or sequentially in a dynamic fashion and in cycles for better treatment of persistent bacterial infections. This Yin–Yang model can also be applied to other microbes, such as fungi, parasites, viruses, and their infections and even cancer and the respective treatments of infections and cancer.[19]

of recovery or reversion and cause relapse or prolonged infections with symptoms. The Yin–Yang model can also be applied to genetic drug resistance (Yang resistance) in growing bacteria where bacteria grow in the presence of antibiotics due to spontaneous mutations or mobile genetic elements (plasmid or transposon), as well as phenotypic resistance[17,18] or antibiotic tolerance (Yin resistance, non-inheritable), in non-growing persisters due to physiological or epigenetic changes (gene expression, protein or DNA modifications). The two types of resistances may overlap and interconvert. The Yin–Yang model can also be used to explain varying hierarchy or spectrum of latent infections (Yin) and active disease (Yang) at the host level and their respective interconversions.[8] This Yin–Yang model can also be applied to other microbes besides bacteria, such as fungi, parasites, and viruses (viral infected cells), and their infections and even cancer and cancer treatments (see below).[19] A list of studies and observations that support or are consistent with the Yin–Yang model is presented in Table 1.

The Yin–Yang model simplifies and provides a unified model for the complex persister phenomenon and heterogeneity and hierarchy of persisters at the bacterial level and also persistent infections at the host level (see below). In addition, the Yin–Yang model explains the current practice of using isoniazid (INH), a drug only active against growing mycobacteria, for the treatment of latent tuberculosis (TB) infection as well as the current practice of two phase TB therapy where the second phase continues use of INH after the first phase of treatment with four drugs (INH, rifampin, pyrazinamide and ethambutol), which should have killed all growing bacteria already (Table 1). In addition, the Yin–Yang model proposes the use of multiple drugs targeting different bacterial populations, both persisters (Yin) and growing bacteria (Yang) for improved therapeutic effect. (See Figure 1 for more details.)

Persisters have been divided into two groups. Type I persisters (non-growing persisters formed in response to external triggers such as starvation) exit slowly from the stationary phase and do not grow in numbers from log phase to stationary phase. Type II persisters (slowly growing) are formed by phenotypic switching in the absence of external triggers and can switch back to normal phenotype and grow in numbers during the growth phase.[20] The classification of type I and type II persisters is useful in characterizing persisters; however, it is worth noting that persisters are much more heterogeneous than the terms type I and type II suggest because either type I or type II persisters themselves again consist of different heterogeneous persisters within each category and the two types of persisters may interconvert as described in the Yin-Yang model.

Persister phenomenon is present in virtually all bacterial species, but the degree of persistence may vary among species as well as within species.[21] In addition, persisters can adopt varying sizes and shapes from regular morphology to altered morphologies (granular or coccoid) as found in old cultures, biofilms and L-form bacteria.[8,18,22,23] L-form bacteria are atypical, pleomorphic cell wall-deficient forms that are formed as part of the life cycle of stressed bacteria and have been implicated in persistent infections.[23] Similarities between L-form bacteria, biofilm bacteria and persisters have been found[22] and are discussed below (see the section on 'MECHANISMS OF PERSISTER FORMATION AND SURVIVAL').

Persisters and multidrug tolerance

Persisters show tolerance to various bactericidal antibiotics, a property called multidrug tolerance (MDT). It was proposed that MDT in persisters is due to the prevention of 'corruption' of drug targets by antibiotics in persister bacteria,[13] but there is no evidence to support this hypothesis, and detailed mechanisms involved in MDT are not well understood. Recent studies have shown that there are multiple mechanisms of MDT. These mechanisms include reduced production in persisters of reactive oxygen species (ROS) influenced by the levels of antioxidant enzymes,[24,25] inhibition of macromolecule synthesis by toxin–antitoxin (TA) modules,[26] increased suppression of cellular metabolism mediated by PhoU[4] and the presence of defects in trans-translation pathway that confer a broad defect in MDT.[27] Decreased antibiotic uptake was recently shown to be involved in drug tolerance to fluoroquinolones, rifampin and linezolid in nutrient starved *Mycobacterium tuberculosis*.[28] It remains to be seen whether reduced permeability to antibiotics is also found in other bacterial species as a mechanism for MDT in persisters. Although antibiotic tolerance in persisters is thought to be phenotypic, it is possible that under some conditions, antibiotic tolerant persisters may acquire mutations and develop genetic resistance. Similarly, a genetically antibiotic resistant mutant (Yang resistance) could also develop persisters with tolerance (Yin resistance); thus, genetic resistance and tolerance may interconvert and overlap.[8]

Stress and persisters

Because persisters are tolerant to not only antibiotics but also other stresses, susceptibility to stresses of mutants is often tested as part of

Table 1 Studies and observations that support or are consistent with the Yin–Yang model (see Figure 1)

Setting	Organisms	References
Inclusion of pyrazinamide that kills persisters with other drugs that kill growing bacilli shortens TB treatment in mice and humans	*M. tuberculosis*	78,106–108
Two phases of TB therapy where the first phase involves a combination of INH, RIF, EMB and PZA followed by the second phase of only INH and RIF. INH is a drug that only kills growing bacteria and its inclusion in the second phase of treatment is to kill the 'reverters' from the persisters not killed by the first phase treatment	*M. tuberculosis*	8,109
Use of isoniazid, a drug that is only active for growing bacteria, for treatment of LTBI; during LTBI, there are growing TB bacteria (reverters) that are susceptible to INH	*M. tuberculosis*	8,109
Spectrum or varying levels of persistence during latent TB infection and treatment	*M. tuberculosis*	51,106
Rapidly growing bacteria can give rise to persisters, whereas stationary phase bacteria can have cryptic growth	*E. coli*	12,110
Heterogeneity of persisters as demonstrated by varying antibiotic exposure times: 'shallow' persisters and 'deep' persisters	*E. coli*	4,9
Cancer stem cell drug candidates used in combination with current cancer drugs improve cancer treatment in mice	Breast cancer	111–113

Abbreviations: EMB, ethambutol; LTBI, latent TB infection.

the persister phenotypes in evaluating persister-defective mutants. For example, *phoU* and *sucB* mutants with defects in persisters are highly susceptible to not only antibiotics but also a variety of stresses.[4,9] On the other hand, stresses can slow and inhibit bacterial growth, resulting in lower metabolic status and facilitates persister formation. Nutrient (amino acid or carbon) depletion has been shown to induce drug tolerant persisters.[29] The carbon starvation mediated persister formation is mediated through activation of the ppGpp-SpoT metabolic TA module, which then leads to inhibition of DNA-negative supercoiling, a process that is affected by FIS, IHF, HU and SeqA DNA-binding proteins that participate in ppGpp-dependent persister formation through modulating DNA negative supercoiling.[30,31] However, the persisters induced by transient nutrient depletion seem to lack the sustainable, multidrug-tolerant phenotype of persisters in the stationary-phase population.[29] Heat, acidic pH and oxidative stresses have been shown to induce persister formation.[32,33] Notably, bacterial persisters can tolerate antibiotics by reducing production of hydroxyl radicals.[24] Although defects in the stringent response genes *relA* and *spoT* are known to cause decreased antibiotic tolerance,[31] this phenotype was recently shown to be mediated through reduced production of the antioxidant defense enzymes superoxide dismutase and catalase.[25,34] Furthermore, inactivation of enzymes involved in hydrogen sulfide (H_2S) production in various bacteria rendered the bacteria highly sensitive to a variety of antibiotics due to loss of H_2S antagonism of the reactive oxygen species induced by antibiotics.[35] Like H_2S, NO has also been shown to induce antibiotic tolerance through antioxidative defense.[36] Low concentrations of antibiotics, such as ciprofloxacin, which presumably causes reactive oxygen production and reduced membrane potential via toxin TisB, could induce persister formation.[37] More recently, antibiotics, such as the RNA synthesis inhibitor rifampin, protein synthesis inhibitor tetracycline and energy inhibitor CCCP, were shown to induce persister formation and enrich the proportion of persisters in cultures.[38] Arrested protein synthesis caused by the above diverse stresses seems to be involved in persister formation.[38]

Persister assays and models

The current persister assays consist of exposing bacterial cultures or cells to bactericidal antibiotics (cell wall inhibitors, aminoglycosides or quinolones) for a short period of time (usually 2–6 h) and then scoring the number of surviving bacteria by colony-forming unit assay.[39,40] Some studies added antibiotics directly to stationary phase cultures, which has more persisters not killed, whereas others resuspended or diluted stationary phase cultures in fresh medium containing antibiotics,[41] which typically leads to fewer persisters due to elevated metabolic activity of stationary phase bacteria being resuspended in fresh medium. These different conditions affect persister counts. In addition, the type of antibiotics, antibiotic exposure time, antibiotic concentrations, age of cultures, aeration and culture media all affect persister numbers.[4,9,41] The recovery time after antibiotic exposure may vary among persister cells.[2,10] An automated method, ScanLag, was recently developed to detect delayed growth of persisters and is useful for measuring the slow recovery of persisters.[42] There is a tendency in the field toward frequently using short antibiotic exposure times of no more than 6–8 h in persister assays. It must be emphasized that while a short exposure time to antibiotics is sufficient for demonstrating the presence of persisters, it may not be sufficient to demonstrate persister defects in some mutants that are obvious only after prolonged antibiotic exposure.[4,9] In fact, the original studies by Hobby and Bigger used penicillin exposure times of 24 h or 48 h and even up to 3–11 days.[1,2] If one understands the enormous heterogeneity of persisters, as expressed in 'shallow' and 'deep' persisters[9] and best captured in the Yin–Yang model,[5,8] one may not need to be so dogmatic about sticking to short antibiotic exposure times in persister assays. In addition, although the original persister phenomenon was demonstrated with bactericidal antibiotics, stress conditions have also been used as an equivalent to antibiotics in persister studies.[4,9,43,44] It is likely that there is overlap between antibiotic persisters and stress persisters despite individuality or heterogeneity and specificity of persisters to particular conditions. This can be addressed using single-cell techniques such as utilizing a microfluidic device (see below). In addition, one has to determine which persister model among different models to use and whether one persister model is more relevant than others in persister studies. Finally, it must be realized that *in vitro* persisters are not the same as *in vivo* persisters due to differences in the environments that the bacteria reside in and the presence or absence of antibiotic exposure. Thus, a drug that can kill all *in vitro* persisters is not guaranteed to do so *in vivo*. Nevertheless, the *in vitro* persisters may share some common features of *in vivo* persisters and *in vitro* persister models should still have value in persister studies as surrogates of *in vivo* persisters. Even *in vivo*, persisters are not all the same and are subject to hierarchy and heterogeneity of persisters as expressed in the Yin–Yang model (Figure 1).

Persisters and single-cell analysis

Although persister cells were found to be dormant or non-growing at the population level in the 1940s, the presence of single persister bacteria tolerant to antibiotics was demonstrated convincingly only recently, using a single-cell microfluidic device.[20] There is recent interest in the use of single-cell techniques for the study of persisters.[20,45] The single-cell techniques are powerful for demonstrating tolerance to cidal antibiotics in a single persister,[20] yet so far no transcriptomic or

proteomic data are available for single persister cells due to the lack of sensitivity of the current methods. With increasing appreciation of the heterogeneity of persisters,[5,7,8] the single-cell technique also faces some challenges as to which persister to study and whether the persister cell obtained in one *in vitro* system would be representative of other persisters in the population *in vitro* and persister cells *in vivo*. Recently, microfluidics studies revealed that *Mycobacterium smegmatis* cells expressing lower levels of KatG expression were tolerant to INH and grew in the presence of INH.[45] INH is a prodrug that needs to be activated by the KatG enzyme, mutations of which cause INH resistance.[46] It was proposed that stochastic expression of KatG leading to various bacterial populations expressing different amounts of KatG can lead to different INH tolerant persister populations.[45] Although this *in vitro* model explains varying susceptibility or tolerance to INH as a function of the level of KatG expression in an artificial system, this may not be used as an argument against persisters being non-growing or dormant. It remains to be determined if this is a relevant persister mechanism for generation of real INH persisters *in vivo* or even *in vitro* in stationary phase cultures.

PERSISTERS, LATENT INFECTIONS AND PATHOGENESIS

Persisters pose significant challenges for the treatment of many chronic and persistent bacterial infections such as TB,[8] Lyme disease[47] and urinary tract infections (Table 2). Persisters underlie latent infections, chronic and recurrent infections, biofilm infections and lengthy therapy of certain bacterial infections, such as TB, and post-treatment persistence and relapse.[8,13,18,48,49] While the most attention has been given to genetic drug resistance either in bacteria, viruses or even cancer, persistence or tolerance to antibiotics (Yin resistance) is equally important to, if not more important than, genetic drug resistance (Yang resistance) because prolonged and repeated treatment of persistent infections may lead to genetic drug resistance, which could occur during TB treatment.

Persistent and latent infections are more complex than previously thought and are found to be of varying hierarchy[50] and in continuous spectrum[51] and can be expressed in the Yin–Yang model (Figure 1).[8] Persistent or latent infections can be pre-antibiotic persistent or post-antibiotic persistent. Pre-antibiotic persistence that is formed under the pressure of the host immune responses refers to initial latent infection before the development of active disease and antibiotic treatment, whereas post-antibiotic persistence refers to the presence of persisters that survive antibiotic treatment and can relapse after treatment. Pre-antibiotic persistence may not be the same as post-antibiotic persistence, which may be more similar to 'deep' persistence. In addition, microbial variants with increased persistence or antibiotic tolerance may develop during treatment as observed in chronic *Pseudomonas aeruginosa* infection in cystic fibrosis patients,[52] but the molecular basis involved is unclear.

Persistence seems to be a widespread phenomenon. However, different bacterial species seem to have different capacities for persistence *in vitro* and *in vivo* such that bacterial infections have varying degrees of difficulty to treat or cure (Table 2). For example, *Streptococcus pneumoniae* seems to have poor ability to form persisters such that its cure by a single antibiotic can be achieved readily in a week or two. In addition, immune clearance of a small number of residual *S. pneumonia* seems effective, so there is usually no relapse after antibiotic treatment. In contrast, some bacterial species, such as *M. tuberculosis*, cause a chronic persistent infection that takes at least 6 months to cure while the immune system seems to be less adequate to clear residual persisters left over from chemotherapy. More recently, *Borrelia burgdorferi* has been demonstrated to have a persistence problem despite antibiotic treatment using mouse and monkey models,[53,54] which may provide some explanation for persisting chronic Lyme disease observed in some patients.[47] In addition to bacterial factors that vary in persistence, the host susceptibilities that vary among individuals play a role in the degree of persistence during infection as well. These variations at the levels of bacterial persistence and host defense mechanisms can have implications in treatment of bacterial infections and might explain why some individuals develop chronic disease and relapse after treatment, whereas others seem to have a stable cure. A variety of conditions, such as host immune and hormonal factors, physical and psychological stresses, and co-infections, such as HIV, measles and mixed bacterial infections, might cause relapse or reactivation of latent infections.

It is possible that not all bacterial cells of a given pathogenic species can cause successful infections. We hypothesize that 'seeding' with persisters or mother cells (dormant cells where heterogeneous persisters are derived) may be critical for successful establishment of infection and disease. In addition to the metabolic status of the bacterial cells that enter the host, the heterogeneity of host phagocytes might also influence the outcome of infection. Thus, interactions of the heterogeneous nature of populations of bacteria, such as *M. tuberculosis*, and of the macrophages that ingest them might cause a diverse range of possible outcomes. These outcomes include unsuccessful infection, successful infection with a transient immune response (lost after some time due to bacterial clearance), successful infection with a stable prolonged immune response and successful infection with an

Table 2 Diseases with known bacterial persistence problems

Disease	Pathogen	Treatment
Tuberculosis	*M. tuberculosis*	Isoniazid, rifampin, pyrazinamide, ethambutol
Syphilis	*Treponema pallidum*	Penicillins, doxycycline, macrolide
Lyme disease	*Borrelia burgdorferi*	Doxycycline, amoxicillin
Urinary tract infections	*E. coli*, *Enterococcus*, *Pseudomonas aeruginosa*, *Chlamydia*, *Mycoplasma genitalium*	Trimethoprim, amoxicillin, nitrofurantoin, quinolones, doxycycline, macrolide
Peptic ulcer	*Helicobacter pylori*	Amoxicillin, clarithromycin, metronidazole, omeprazole, doxycycline, bismuth
Bacteremia/sepsis	*Staphylococcus aureus*, *Group B Streptococcus*	Various antibiotics
Endocarditis	*Streptococcus*, *Staphylococcus*, *Enterococcus*	Penicillins, vancomycin
Otitis media	*S. pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*	Amoxicillin, azithromycin
Brucellosis	*Brucella abortus*	Doxycycline, rifampin
Biofilm infections, periodontitis, prosthetic device infections	Various pathogens	Refractory to antibiotic treatment

immune response and disease pathology. This hypothesis needs to be addressed with animal models in future studies. At the level of granuloma lesions, there might be a varying degree of heterogeneous granulomatous tissue correlating to the degree of inflammation, ranging from quiescent granulomas with low inflammation to more active and dynamic granulomas with more active inflammation even in the same lungs, and over time. Recently, it has been shown that there are varying degrees of latent TB infection, ranging from nearly active TB to a latent state with a remote chance of reactivation.[51]

At the host level, it is possible that infection of host stem cells (or stem-like cells or progenitor cells, including quiescent resting memory cells) by pathogens, such as the intracellular bacteria *M. tuberculosis* and *Brucella abortus*, and viruses, such as HIV[55] and HBV,[56] might contribute to increased persistence problems and protracted or chronic disease courses due to the longevity of the stem cells. It is of interest to note that infection with *M. leprae*,[57] which causes chronic leprosy, could reprogram the host cell into a stem cell-like phenotype that survives a long time, though it may not be easy to distinguish if the infected cell is stem cell-like before or after infection. More recently, it was shown that *M. tuberculosis* could reside in bone marrow $CD271^+$/$CD45^-$ mesenchymal stem cells, which could provide a niche for dormant infection.[58] It remains to be seen if the chronicity of infections by certain pathogens, such as mycobacteria, could involve host stem cells as a niche for perpetuation of the infection.

MECHANISMS OF PERSISTER FORMATION AND SURVIVAL

Mechanisms of persister formation are not well understood as persisters are elusive, small in number, heterogeneous, and transient and can change with environment, which poses significant challenges to their study. Epigenetic factors can promote bacterial persister formation through bistable gene expression,[59] mediated through stochastic or induced expression of persister related genes,[60] or through changes in DNA modifications or signaling protein modifications. Thus, permutations at the levels of expression of multiple persister genes (Table 3), regulatory RNA, modifications of DNA and post-translational modifications of proteins could produce enormous diversity and heterogeneity of persisters as expressed in the Yin–Yang model (Figure 1). Although senescence or aging has been proposed as a persister mechanism,[61] aging itself can hardly be a mechanism of persisters as aging must in turn be acting through certain cellular processes, which could involve persister mechanisms. Although various persister genes have been identified (Table 3), what and how cells sense to form persisters remain unclear.

The approaches used to identify persister genes are worth mentioning. Although persisters are caused by epigenetic changes, mutagenesis has been traditionally used to isolate genes involved in persister formation and has led to identification of a range of persister related genes, such as *hipA*, *relA*, *phoU*, *sucB* and *ubiF*, just as sporulation is an epigenetic trait that has become reasonably well understood using the mutagenesis approach. The mutagenesis approach has been used to identify persister-related genes whose mutations caused either reduced persistence or increased persistence. Mutations that cause decreased persistence include *relA*,[31] *phoU*,[4] *sucB* and *ubiF*.[9] Mutations that cause increased persistence map to the following genes: *hipA* encoding toxin,[3] *metG* encoding methionyl-tRNA synthetase, *tktA* encoding transketolase A and *glpD* encoding glycerol-3-phosphate dehydrogenase.[62] In an overexpression study, *glpD* and *glpABC* encoding glycerol-3-phosphate dehydrogenase and *plsB* encoding glycerol-3-phosphate acyltransferase were found to confer increased persistence.[63] It is intriguing that *glpD* had opposite phenotypes in the two different studies.[62,63] However, the mutagenesis approach is only useful for identifying non-essential dominant genes that have a major effect on the phenotype and is less useful for identifying a phenotype that is determined by multiple genes of minor effect. The fact that certain mutagenesis screens to identify persister genes did not provide much insight into persister mechanisms [11,13] does not invalidate this approach to studying persisters. Factors that might have contributed to failure to identify persister genes by the transposon mutant approach might include screening a partial mutant library, short antibiotic exposure and aeration during antibiotic exposure. The duration of antibiotic exposure in the mutant screen is critical. A short exposure of 6 h with ofloxacin was used to screen the *Escherichia coli* KEIO mutant library and identified many genes involved in stress responses and global regulation with minor or 'shallow' persister phenotypes.[39] It is unrealistic to expect a complete loss of persisters (a 'persisterless' phenotype) by a mutant in a screen with a brief antibiotic exposure of a few hours, especially when using stationary phase cultures. A longer antibiotic exposure or higher antibiotic concentrations may be needed for identification of true or 'deep' persister genes and, indeed, has led to the discovery of *phoU*,[4] *sucB* and *ubiF*[9] as persister genes. It is likely that different persister genes will be identified at different antibiotic exposure times. However, a potential limitation of the use of a deletion mutant library for persister

Table 3 Persister mechanisms in bacteria

Persister pathways	Genes involved	Mechanisms/features	References
Toxin–antitoxin modules	*hipBA, relBE, mazEF, tisAB, mqsR, hhA, hokA, cspD, pasT*	Toxin–antitoxin modules inhibit protein or nucleic acid synthesis; Lon protease can degrade the antitoxin to regulate persister formation	3,12,37,40,69,114,115
Alternative energy production	*sucB, ubiF, glpD, plsB, tgs1*	Provision of energy under stress conditions	9,63,116
Stringent response	*relA, dksA*	ppGpp synthesized by RelA inhibits RNA synthesis	31,34,68,117
SOS response/DNA repair and protection of DNA	*lexA, recA, recB, xerC, xerD, dps*	Repair of DNA damage caused by ROS	29,118–120
Antioxidant defense H_2S, NO	Superoxide dismutase, catalase	Removal of ROS and hydroxyl radical	24,33,35,36
Enhanced efflux or transporter activity	Various	Removal of toxic substances or antibiotic buildup, underlying tolerance to antibiotics and stresses	33,121,122
Phosphate metabolism	*phoU*	PhoU is a negative regulator of phosphate uptake, mutant has dramatic defect in persister phenotype; shutdown of metabolic activity	4,67,123
Trans-translation	*ssrA, smpB, rpsA*	Degradation of toxic proteins and mRNA and recycling of ribosomes	27,81
Signaling pathways	*comE/comC; tnaA, oxyR, flu, pspBC*	Quorum sensing peptide or homoserine lactone or indole, acting through TA or antioxidant defense OxyR and phage-shock pathways	32,124,125

mutant screens is that compensatory mutations could mask the persister defective phenotype, which may lead to an inability to identify critical persister genes. Although microarray analysis has been used for profiling persister related genes,[64,65] the data were obtained mostly on heterogeneous populations, which could mask the signals in individual or single persister cells. In addition, the genes involved in persistence are likely to vary according to the specific environment or models used in the study. These are the challenges facing studies aimed to identify persister genes.

Although different bacterial species may differ in terms of their ability to form persisters, they share many common features and mechanisms. It is increasingly clear that multiple mechanisms of varying hierarchy and importance are involved in persister formation in different models of persistence (Table 3). Our comparative analyses of the pathways involved in persister formation and survival between *E. coli* and *M. tuberculosis*[8] indicate that while persister genes and pathways may vary, the overall persister mechanisms and pathways in different bacterial species are largely conserved (convergent evolution) (Table 3). In addition, the genes and pathways in persisters and biofilm bacteria and L-form bacteria have been found to overlap and share significant similarities,[22] which include SOS response and DNA repair, iron homeostasis, signaling, efflux/transporter, envelope/membrane stress, energy production, phosphate metabolism, sulfur metabolism, signaling, phage shock proteins and protein degradation (protease and trans-translation). These findings suggest that biofilm bacteria, L-form bacteria and persisters are related entities that share common mechanisms.

Given the recent advances in understanding persister mechanisms, it remains to be seen whether the *in vitro* identified persister mechanisms (Table 3) are operative and valid for *in vivo* persisters. Some persister genes, such as *phoU* and *relA*, that have been shown to be a persister gene *in vitro*[4,31] are also involved in virulence[66] and persistence *in vivo*.[67,68] Deletion of TA module PasTI, but not other TA modules, such as HipBA and HigBA, in pathogenic *E. coli*, was shown to have reduced persister formation and decreased virulence in mice.[69]

PERSISTERS, L-FORM AND BIOFILM BACTERIA, AND CANCER STEM CELLS

There are significant parallels between bacterial persisters and cancer stem cells. In cancer, there is a situation analogous and equivalent to bacterial persisters, termed 'cancer stem cells'. Cancer stem cells are defined as 'a small subset of cancer cells within a cancer that constitutes a reservoir of self-sustaining cells with the exclusive ability to self-renew and to cause the heterogeneous lineages of cancer cells that comprise the tumor.[70] It was proposed that cancer stem cells resemble bacterial persister cells in 2007 (http://forms.asm.org/microbe/index.asp?bid551533),[71] based on the common pathways between bacterial persisters, biofilm and L-form bacteria (cell wall-defective variants formed under cell membrane stress)[22,23,72] and cancer stem cells.[19,73] In *E. coli*, L-form bacteria, which can be considered as a type of deep or true persisters (mother cells), occur at the frequency of 10^4–10^5 cells,[22] which is about two orders of magnitude less frequent than persisters. Like persister cells, cancer stem cells are also quite heterogeneous and resist chemotherapy drugs and stresses and cause relapse and metastasis.[19,74] There is significant recent interest in the analogy between bacterial persisters and cancer stem cells.[75–77] The above analyses[19,22,73] revealed that although the genes involved in the common pathways between bacterial persisters, L-form and biofilm bacteria, and cancer stem cells do not show significant homology, they have similar functions. Such parallels in bacterial persisters and cancer stem cells may not only help to shed light on their mechanisms via convergent evolution but also may allow common treatment strategies to be developed for more effective treatment of persistent infections and cancer in the future (see below).

TAMING PERSISTERS: TREATMENT STRATEGIES

While different bacterial infections seem to have different capacities for persistence and varying degrees of difficulty for treatment, their cure relies on the combined action of antibiotics and the host immune system. In addition, the type of drugs and the status of the target cells affect treatment outcome. Here it may be instructive to examine in some detail the interesting example of the unique TB persister drug PZA, which may shed light on the treatment of persistent bacterial infections in general and even cancers. PZA plays a key role in shortening TB therapy from 9–12 months to 6 months by killing a subpopulation of persisters not killed by other TB drugs (Figure 1).[78] PZA is an unconventional and paradoxical drug that acts only on non-growing persisters at acidic pH.[78,79] Unlike common antibiotics that act on growing bacteria, PZA is completely dissimilar in that it has no activity against growing *M. tuberculosis* bacteria.[78] In contrast to common antibiotics that inhibit cell wall, protein, and nucleic acid synthesis and are active only against growing bacteria, PZA inhibits energy production[80] and the trans-translation process, which recycles ribosomes and degrades toxic protein buildup under stress,[81] and perhaps coenzyme A synthesis[82], which is required for survival of *M. tuberculosis* persisters. It is these unique properties of PZA that are critical for killing persisters and shortening TB therapy. It is of interest to note that PZA also inhibits the quiescent malaria parasite in the mouse model[83] and is also active against *E. coli* ampicillin tolerant persisters.[84] Although there is considerable recent interest in developing antibiotics targeting persisters,[13,85,86] PZA is the only prototype persister drug so far that has been shown to improve the treatment of a persistent infection. Nevertheless, PZA validates an important principle that drugs targeting dormant persisters, when used in combination with drugs that target growing organisms, are critical for shortening the treatment. The story of PZA has important implications for developing future antibiotics and cancer drugs that target persisters and cancer stem cells to improve treatment of both persistent infections and cancers and perhaps even latent viral infections, such as HIV and HBV, which hide in quiescent stem-like cells, and also persistent parasites or fungi.

In addition to the insights from the above example, several approaches should be explored to better control persisters. One approach would be to directly target persisters with drugs, but unfortunately all current antibiotics, except the TB drug PZA, are predominantly active against growing bacteria. Current antibiotics generally have no activity against persisters because these types of cells were not used during the screening. There is currently increasing interest in developing new drugs active against bacterial persisters.[7,8,49,87] Some candidate compounds that are active against persisters[8] have been identified and, if they pass the safety and efficacy phase, are expected to be used together with current antibiotics or drugs for improved treatment based on the common principle of targeting both growing bacteria or cells and non-growing persisters.[5,19] This is exemplified in the case of INH (which kills growing bacteria) and PZA (which kills persisters) for TB treatment (Figure 1). However, it is preferable that the drugs in combination interfere with different pathways in the cells and kill different cell populations to optimize the potential for killing of persisters.

A second approach would be to 'wake up' or alter the metabolic status of persisters,[8,18] so they respond to antibiotic treatment.

Although resuscitation factors have been found for bacteria,[16,88] they have not been used therapeutically in animal models to demonstrate feasibility. Recently, metabolites, such as glucose, glycerol and relatively less efficient carbon sources (mannose, fructose, sorbitol, pyruvate, lactate and acetate), and nucleotides, such as thymidine, uridine and inosine, have been shown to potentiate activity of aminoglycoside activity for persisters *in vitro*.[89] Such an approach needs to be validated in animal or human studies in the future.

A third approach would be to enhance the activity of current antibiotics by certain agents to kill some persister cells.[90–92] For example, aspirin, ibuprofen and iron have been shown to enhance the activity of the persister drug PZA against *M. tuberculosis*.[90,91] In addition, sugar mannitol can enhance the killing activity of persisters by aminoglycoside antibiotics through stimulating the proton motive force needed for increased uptake of the antibiotic in the mouse model of urinary tract infection.[92] However, it is unclear whether mannitol works through its diuretic effect to wash off the bacteria more effectively by increasing the amount of urine and/or through its effect on enhancing the uptake of aminoglycoside. In addition, this is a highly specific case, and the sugar only increases the activity of aminoglycosides but not other antibiotics. Furthermore, it remains to be seen whether the use of mannitol is effective in patients. A related approach to enhancing the effectiveness of the existing antibiotics in killing persisters is to increase ROS production.[93] Recently, it has been shown that silver, which produces ROS, enhanced the activity of vancomycin, improving the treatment of bacterial infections in mice.[94] In addition, 3-[4-(4-methoxyphenyl)piperazin-1-yl]piperidin-4-yl biphenyl-4-carboxylate (C10)[95] and (Z)-4-bromo-5-(bromomethylene)-3-methylfuran-2(5H)-one (BF8)[96] were found to convert antibiotic tolerant persisters to an antibiotic sensitive phenotype. It remains to be determined how the compounds work and whether they can be used to resuscitate persisters for improved treatment of persistent infections.

The fourth approach would be to harness the host immune system to control persisters and cancer stem cells through enhancing innate and acquired immunity in the form of immune-modulating cytokines or immunotherapeutic vaccines that encompass antigens from both growing cells and non-growing cells (persisters and cancer stem cells). For example, inclusion of antigens from both growing bacteria (Antigen 85 and ESAT-6) and dormancy antigen Rv2660c or HspX from *M. tuberculosis* could enhance vaccine efficacy in prophylactic and therapeutic vaccines in animal models.[97] Combined immunotherapy with chemotherapy for persisters should also be explored for improved treatment.[98,99]

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Despite significant progress in our understanding of persisters in recent years, much remains to be learned about the biology of persisters. Classical genetic approaches have identified multiple genes and pathways that are involved in persister formation or survival. However, there are some limitations with the classical genetic mutant approach due to problems of compensatory mutations and with the reductionist approach of looking at one gene at a time. With the application of the 'omics' (transcriptome, proteome, metabolome, epigenome) and next-generation sequencing techniques including Tn-seq,[100] new knowledge about persisters will undoubtedly be gained in the near future. Networks and systems biology approaches remain to be applied to the study of persister mechanisms. It is not enough to say that the whole is more than its parts in terms of persister mechanisms. More importantly, how different components interact in a dynamic manner in the context of systems biology to cause complex persister phenotypes needs to be addressed in the future. An even more crucial challenge is to understand what these complex data mean and whether useful intervention or treatment strategies can be derived from them. In addition, significant technical hurdles exist when applying the 'omics' tools to single or rare true persister cells or mother cells due to lack of sufficiently sensitive techniques. For example, current single-cell techniques cannot yet identify the transcriptomic or proteomic profiles of an individual persister cell. In addition, future studies may need to explore new dimensions of persister mechanisms, including studying possible roles of bio-electromagnetic fields and information flow ('Qi' or flow of energy or life force) in cellular circuits that maintain viability of persisters in the context of thermodynamics as in dissipative systems. It is in this context that the relationships between stress, cell death, aging, persistence and longevity and the nature of life need to be investigated in the future.

While most persister mechanistic studies have been performed mainly with *E. coli*, it is important to study the persister mechanisms in other bacterial pathogens. As an evolutionarily useful strategy to survive harmful stresses in the environment, the persister phenomenon occurs not only in bacteria but also in all life forms (kingdoms). For example, persister phenomenon has been found in fungi,[101] parasites,[102] cancer stem cells[19] and viral infected host cells.[55] It would be of interest to compare and contrast the common mechanisms among bacteria, fungi, parasites[102] and viral (HIV,[55] HBV, HPV) infected host cells and cancer stem cells.[19]

It would be of interest to study the latent forms of the disease (i.e., latent infections) rather than just the advanced and complicated forms of the disease. Future studies by ecological approaches need to examine the microenvironment of persisters and assess the environmental factors, as well as host factors (including role of host microbiota), that affect reactivation, progression and outcome of the disease. In addition, it would be of interest to develop more sensitive diagnostic tools to detect dormant persister organisms in clinical specimens and in affected tissues. Moreover, it will be necessary to identify immune mechanisms that control latent infections. Such information will be useful for developing interventions based on altering the microenvironment needed for survival of persisters and developing immunotherapeutic vaccines for their effective control. It is important to understand why some individuals are not cured while others are cured. Future investigations are needed to understand why some individuals seem to have chronic persistent and recurrent infections, whereas other individuals are cured by standard treatment in the context of varying degrees of host susceptibilities (defined in a broad sense not necessarily restricted by genetic factors) and bacterial persistence.

It is important to establish more relevant models of persisters or persistence for mechanistic studies that are representative of *in vivo* situations, as well as developing drugs that kill *in vivo* persisters and improve treatment. It would be quite challenging to develop persister drugs as one ponders which model to use for drug screens, considering the diverse and variable nature of persisters as expressed in the Yin–Yang model (Figure 1).[8] The above problems with bacterial persisters, also apply to cancer stem cells,[19,103] and will be a major stumbling block for both fields and a major topic of interest for the future. The current *in vitro* models of persisters or cancer stem cells may have significant limitations and it remains to be seen if the data obtained *in vitro* can be validated *in vivo* in animal models or patients.

There are currently significant debates, as well as interest, about persister mechanisms and drugs. To capture the current status of the field, it may be fitting to end the article with the parable about the blind men and the elephant. The elephant, which is analogous to

persisters or cancer stem cells, is described as a snake, a spear, a fan, a tree, a wall and a rope by blind men touching different parts of the elephant, which represent different models and pathways of persisters or cancer stem cells and are only partially right. This partial knowledge, which largely results from the limitations of current methodologies, is not perfect and is an intermediate state of knowledge that is useful and acceptable with reservation. The ultimate test of this partial knowledge will be whether we can devise useful drugs and therapeutic strategies targeting persisters for improved treatment in the future. There is a convergence of interest in both the persister field and the cancer stem cell field to develop new drugs targeting the quiescent forms ('Yin') (i.e., persisters)[8,49,85,87] of cancer stem cells for improved treatment.[19,104,105] The identified pathways in bacterial persisters could serve as potential targets for development of new persister drugs. From the prototype persister drug PZA, one may see the future of antibiotic and even cancer drug development. Future studies are needed to test whether drugs analogous to PZA that target persisters and cancer stem cells can improve treatment of persistent infections and cancers.

ACKNOWLEDGEMENTS

The support from NIH AI099512, Lyme Research Alliance and lymedisease.org is gratefully acknowledged. I thank Peng Cui (Huashan Hospital, Fudan University, China) for help with drawing Figure 1.

1 Hobby GL, Meyer K, Chaffee E. Observations on the mechanism of action of penicillin. *Proc Soc Exp Biol NY* 1942; **50**: 281–285.
2 Bigger JW. Treatment of staphylococcal infections with penicillin by intermittent sterilisation. *Lancet* 1944; **244**: 497–500.
3 Moyed HS, Bertrand KP. hipA, a newly recognized gene of *Escherichia coli* K-12 that affects frequency of persistence after inhibition of murein synthesis. *J Bacteriol* 1983; **155**: 768–775.
4 Li Y, Zhang Y. PhoU is a persistence switch involved in persister formation and tolerance to multiple antibiotics and stresses in *Escherichia coli*. *Antimicrob Agents Chemother* 2007; **51**: 2092–2099.
5 Zhang Y. Advances in the treatment of tuberculosis. *Clin Pharmacol Ther* 2007; **82**: 595–600.
6 Gefen O, Balaban NQ. The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. *FEMS Microbiol Rev* 2009; **33**: 704–717.
7 Allison KR, Brynildsen MP, Collins JJ. Heterogeneous bacterial persisters and engineering approaches to eliminate them. *Curr Opin Microbiol* 2011; **14**: 593–598.
8 Zhang Y, Yew WW, Barer MR. Targeting persisters for tuberculosis control. *Antimicrob Agents Chemother* 2012; **56**: 2223–2230.
9 Ma C, Sim S, Shi W, Du L, Xing D, Zhang Y. Energy production genes sucB and ubiF are involved in persister survival and tolerance to multiple antibiotics and stresses in *Escherichia coli*. *FEMS Microbiol Lett* 2010; **303**: 33–40.
10 Joers A, Kaldalu N, Tenson T. The frequency of persisters in *Escherichia coli* reflects the kinetics of awakening from dormancy. *J Bacteriol* 2010; **192**: 3379–3384.
11 Hu Y, Coates AR. Transposon mutagenesis identifies genes which control antimicrobial drug tolerance in stationary-phase *Escherichia coli*. *FEMS Microbiol Lett* 2005; **243**: 117–124.
12 Keren I, Kaldalu N, Spoering A, Wang Y, Lewis K. Persister cells and tolerance to antimicrobials. *FEMS Microbiol Lett* 2004; **230**: 13–18.
13 Lewis K. Persister cells. *Annu Rev Microbiol* 2010; **64**: 357–372.
14 Xu HS, Roberts N, Singleton FL, Attwell RW, Grimes DJ, Colwell RR. Survival and viability of nonculturable *Escherichia coli* and *Vibrio cholerae* in the estuarine and marine environment. *Microbial Ecology* 1982; **8**: 313–323.
15 Hobby GL, Auerbach O, Lenert TF, Small MJ, Comer JV. The late emergence of *M. tuberculosis* in liquid cultures of pulmonary lesions resected from humans. *Am Rev Tuberc* 1954; **70**: 191–218.
16 Zhang Y, Yang Y, Woods A, Cotter RJ, Sun Z. Resuscitation of dormant *Mycobacterium tuberculosis* by phospholipids or specific peptides. *Biochem Biophys Res Commun* 2001; **284**: 542–547.
17 Greenwood D. Phenotypic resistance to antimicrobial agents. *J Antimicrob Chemother* 1985; **15**: 653–655.
18 Zhang Y. Persistent and dormant tubercle bacilli and latent tuberculosis. *Front Biosci* 2004; **9**: 1136–1156.
19 Zhou J, Zhang Y. Cancer stem cells: models, mechanisms and implications for improved treatment. *Cell Cycle* 2008; **7**: 1360–1370.
20 Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. *Science* 2004; **305**: 1622–1625.
21 Stewart B, Rozen DE. Genetic variation for antibiotic persistence in *Escherichia coli*. *Evolution* 2012; **66**: 933–939.
22 Glover WA, Yang Y, Zhang Y. Insights into the molecular basis of L-form formation and survival in *Escherichia coli*. *PLoS ONE* 2009; **4**: e7316.
23 Domingue GJ. Demystifying pleomorphic forms in persistence and expression of disease: are they bacteria, and is peptidoglycan the solution? *Discov Med* 2010; **10**: 234–246.
24 Kim JS, Heo P, Yang TJ, *et al*. Bacterial persisters tolerate antibiotics by not producing hydroxyl radicals. *Biochem Biophys Res Commun* 2011; **413**: 105–110.
25 Bink A, Vandenbosch D, Coenye T, Nelis H, Cammue BP, Thevissen K. Superoxide dismutases are involved in *Candida albicans* biofilm persistence against miconazole. *Antimicrob Agents Chemother* 2011; **55**: 4033–4037.
26 Gerdes K, Maisonneuve E. Bacterial persistence and toxin–antitoxin loci. *Annu Rev Microbiol* 2012; **66**: 103–123.
27 Li J, Ji L, Shi W, Xie J, Zhang Y. Trans-translation mediates tolerance to multiple antibiotics and stresses in *Escherichia coli*. *J Antimicrob Chemother* 2013; **68**: 2477–2481.
28 Sarathy J, Dartois V, Dick T, Gengenbacher M. Reduced drug uptake in phenotypically resistant nutrient-starved nonreplicating *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2013; **57**: 1648–1653.
29 Fung DK, Chan EW, Chin ML, Chan RC. Delineation of a bacterial starvation stress response network which can mediate antibiotic tolerance development. *Antimicrob Agents Chemother* 2010; **54**: 1082–1093.
30 Amato SM, Orman MA, Brynildsen MP. Metabolic control of persister formation in *Escherichia coli*. *Mol Cell* 2013; **50**: 475–487.
31 Korch SB, Henderson TA, Hill TM. Characterization of the hipA7 allele of *Escherichia coli* and evidence that high persistence is governed by (p)ppGpp synthesis. *Mol Microbiol* 2003; **50**: 1199–1213.
32 Leung V, Levesque CM. A stress-inducible quorum-sensing peptide mediates the formation of persister cells with noninherited multidrug tolerance. *J Bacteriol* 2012; **194**: 2265–2274.
33 Wu Y, Vulic M, Keren I, Lewis K. Role of oxidative stress in persister tolerance. *Antimicrob Agents Chemother* 2012; **56**: 4922–4926.
34 Nguyen D, Joshi-Datar A, Lepine F *et al*. Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science* 2011; **334**: 982–986.
35 Shatalin K, Shatalina E, Mironov A, Nudler E. H2S: a universal defense against antibiotics in bacteria. *Science* 2011; **334**: 986–990.
36 Gusarov I, Shatalin K, Starodubtseva M, Nudler E. Endogenous nitric oxide protects bacteria against a wide spectrum of antibiotics. *Science* 2009; **325**: 1380–1384.
37 Dorr T, Vulic M, Lewis K. Ciprofloxacin causes persister formation by inducing the TisB toxin in *Escherichia coli*. *PLoS Biol* 2010; **8**: e1000317.
38 Kwan BW, Valenta JA, Benedik MJ, Wood TK. Arrested protein synthesis increases persister-like cell formation. *Antimicrob Agents Chemother* 2013; **57**: 1468–1473
39 Hansen S, Lewis K, Vulic M. Role of global regulators and nucleotide metabolism in antibiotic tolerance in *Escherichia coli*. *Antimicrob Agents Chemother* 2008; **52**: 2718–2726.
40 Hong SH, Wang X, O'Connor HF, Benedik MJ, Wood TK. Bacterial persistence increases as environmental fitness decreases. *Microb Biotechnol* 2012; **5**: 509–522.
41 Luidalepp H, Joers A, Kaldalu N, Tenson T. Age of inoculum strongly influences persister frequency and can mask effects of mutations implicated in altered persistence. *J Bacteriol* 2011; **193**: 3598–3605.
42 Levin-Reisman I, Gefen O, Fridman O *et al*. Automated imaging with ScanLag reveals previously undetectable bacterial growth phenotypes. *Nat Methods* 2010; **7**: 737–739.
43 Scherrer R, Moyed HS. Conditional impairment of cell division and altered lethality in hipA mutants of *Escherichia coli* K-12. *J Bacteriol* 1988; **170**: 3321–3326.
44 Pearl S, Gabay C, Kishony R, Oppenheim A, Balaban NQ. Nongenetic individuality in the host–phage interaction. *PLoS Biol* 2008; **6**: e120.
45 Wakamoto Y, Dhar N, Chait R, *et al*. Dynamic persistence of antibiotic-stressed mycobacteria. *Science* 2013; **339**: 91–95.
46 Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature* 1992; **358**: 591–593.
47 Stricker RB, Johnson L. The pain of chronic Lyme disease: moving the discourse backward? *FASEB J* 2011; **25**: 4085–4087.
48 McDermott W. Microbial persistence. *Yale J Biol Med* 1958; **30**: 257–291.
49 Fauvart M, de Groote VN, Michiels J. Role of persister cells in chronic infections: clinical relevance and perspectives on anti-persister therapies. *J Med Microbiol* 2011; **60**: 699–709.
50 McDermott W. Microbial persistence. *Harvey Lecture* 1969; **63**: 1–31.
51 Barry CE 3rd, Boshoff HI, Dartois V *et al*. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* 2009; **7**: 845–855.
52 Mulcahy LR, Burns JL, Lory S, Lewis K. Emergence of *Pseudomonas aeruginosa* strains producing high levels of persister cells in patients with cystic fibrosis. *J Bacteriol* 2010; **192**: 6191–6199.
53 Barthold SW, Hodzic E, Imai DM, Feng S, Yang X, Luft BJ. Ineffectiveness of tigecycline against persistent *Borrelia burgdorferi*. *Antimicrob Agents Chemother* 2010; **54**: 643–651.
54 Embers ME, Barthold SW, Borda JT *et al*. Persistence of *Borrelia burgdorferi* in rhesus macaques following antibiotic treatment of disseminated infection. *PLoS ONE* 2012; **7**: e29914.

55 Finzi D, Blankson J, Siliciano JD *et al*. Latent infection of CD4$^+$ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med* 1999; **5**: 512–517.
56 Zoulim F. New insight on hepatitis B virus persistence from the study of intrahepatic viral cccDNA. *J Hepatol* 2005; **42**: 302–308.
57 Masaki T, Qu J, Cholewa-Waclaw J, Burr K, Raaum R, Rambukkana A. Reprogramming adult Schwann cells to stem cell-like cells by leprosy bacilli promotes dissemination of infection. *Cell* 2013; **152**: 51–67.
58 Das B, Kashino SS, Pulu I *et al*. CD271$^+$ bone marrow mesenchymal stem cells may provide a niche for dormant *Mycobacterium tuberculosis*. *Sci Transl Med* 2013; **5**: 170ra13.
59 Veening JW, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* 2008; **62**: 193–210.
60 Rotem E, Loinger A, Ronin I *et al*. Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proc Natl Acad Sci USA* 2010; **107**: 12541–12546.
61 Klapper I, Gilbert P, Ayati BP, Dockery J, Stewart PS. Senescence can explain microbial persistence. *Microbiology* 2007; **153**: 3623–3630.
62 Girgis HS, Harris K, Tavazoie S. Large mutational target size for rapid emergence of bacterial persistence. *Proc Natl Acad Sci USA* 2012; **109**: 12740–12745.
63 Spoering AL, Vulic M, Lewis K. GlpD and PlsB participate in persister cell formation in *Escherichia coli*. *J Bacteriol* 2006; **188**: 5136–5144.
64 Shah D, Zhang Z, Khodursky A, Kaldalu N, Kurg K, Lewis K. Persisters: a distinct physiological state of *E. coli*. *BMC Microbiol* 2006; **6**: 53.
65 Betts J, Lukey P, Robb L, McAdam R, Duncan K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* 2002; **43**: 717–731.
66 Buckles EL, Wang X, Lockatell CV, Johnson DE, Donnenberg MS. PhoU enhances the ability of extraintestinal pathogenic *Escherichia coli* strain CFT073 to colonize the murine urinary tract. *Microbiology* 2006; **152**: 153–160.
67 Shi W, Zhang Y. PhoY2 but not PhoY1 is the PhoU homologue involved in persisters in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* 2010; **65**: 1237–1242.
68 Dahl JL, Kraus CN, Boshoff HI *et al*. The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of *Mycobacterium tuberculosis* in mice. *Proc Natl Acad Sci USA* 2003; **100**: 10026–10031.
69 Norton JP, Mulvey MA. Toxin–antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. *PLoS Pathog* 2012; **8**: e1002954.
70 Clarke MF, Dick JE, Dirks PB *et al*. Cancer stem cells—perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res* 2006; **66**: 9339–9344.
71 Holzman D. *Genetic switch plays key role, converting ordinary to ''persister'' bacteria*. Washington, DC: Microbe Magazine, 2007. Available at http://forms.asm.org/microbe/index.asp?bid551533 (accessed 24 March 2013).
72 Allan EJ, Hoischen C, Gumpert J. Bacterial L-forms. *Adv Appl Microbiol* 2009; **68**: 1–39.
73 Zhou J, Wulfkuhle J, Zhang H *et al*. Activation of the PTEN/mTOR/STAT3 pathway in breast cancer stem-like cells is required for viability and maintenance. *Proc Natl Acad Sci USA* 2007; **104**: 16158–16163.
74 Pietras A. Cancer stem cells in tumor heterogeneity. *Adv Cancer Res* 2011; **112**: 255–281.
75 Glickman MS, Sawyers CL. Converting cancer therapies into cures: lessons from infectious diseases. *Cell* 2012; **148**: 1089–1098.
76 Dawson CC, Intapa C, Jabra-Rizk MA. ''Persisters'': survival at the cellular level. *PLoS Pathog* 2011; **7**: e1002121.
77 Ben-Jacob E, Coffey DS, Levine H. Bacterial survival strategies suggest rethinking cancer cooperativity. *Trends Microbiol* 2012; **20**: 403–410.
78 Zhang Y, Mitchison D. The curious characteristics of pyrazinamide: a review. *Int J Tuberc Lung Dis* 2003; **7**: 6–21.
79 Zhang Y, Permar S, Sun Z. Conditions that may affect the results of susceptibility testing of *Mycobacterium tuberculosis* to pyrazinamide. *J Med Microbiol* 2002; **51**: 42–49.
80 Zhang Y, Wade MM, Scorpio A, Zhang H, Sun Z. Mode of action of pyrazinamide: disruption of *Mycobacterium tuberculosis* membrane transport and energetics by pyrazinoic acid. *J Antimicrob Chemother* 2003; **52**: 790–795.
81 Shi W, Zhang X, Jiang X *et al*. Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. *Science* 2011; **333**: 1630–1632.
82 Zhang S, Chen J, Shi W, Liu W, Zhang WH, Zhang Y. Mutations in panD encoding aspartate decarboxylase are associated with pyrazinamide resistance in *Mycobacterium tuberculosis*.*Emerg Microbes Infect* 2013; **2**: e34.
83 Deye GA, Gettayacamin M, Hansukjariya P, *et al*. Use of a rhesus *Plasmodium cynomolgi* model to screen for anti-hypnozoite activity of pharmaceutical substances. *Am J Trop Med Hyg* 2012; **86**: 931–935.
84 Wade MM, Zhang Y. Effects of weak acids, UV and proton motive force inhibitors on pyrazinamide activity against *Mycobacterium tuberculosis in vitro*. *J Antimicrob Chemother* 2006; **58**: 936–941.
85 Coates AR, Hu Y. Targeting non-multiplying organisms as a way to develop novel antimicrobials. *Trends Pharmacol Sci* 2008; **29**: 143–150.
86 Nathan C. Fresh approaches to anti-infective therapies. *Sci Transl Med* 2012; **4**: 140sr2.
87 Zhang Y. The magic bullets and tuberculosis drug targets. *Annu Rev Pharmacol Toxicol* 2005; **45**: 529–564.
88 Mukamolova GV, Turapov OA, Young DI, Kaprelyants AS, Kell DB, Young M. A family of autocrine growth factors in *Mycobacterium tuberculosis*. *Mol Microbiol* 2002; **46**: 623–635.
89 Orman MA, Brynildsen MP. Establishment of a method to rapidly assay bacterial persister metabolism. *Antimicrob Agents Chemother* 2013; **57**: 4398–4409.
90 Somoskovi A, Wade MM, Sun Z, Zhang Y. Iron enhances the antituberculous activity of pyrazinamide. *J Antimicrob Chemother* 2004; **53**: 192–196.
91 Byrne ST, Denkin SM, Zhang Y. Aspirin and ibuprofen enhance pyrazinamide treatment of murine tuberculosis. *J Antimicrob Chemother* 2007; **59**: 313–316.
92 Allison KR, Brynildsen MP, Collins JJ. Metabolite-enabled eradication of bacterial persisters by aminoglycosides. *Nature* 2011; **473**:216–220.
93 Grant SS, Kaufmann BB, Chand NS, Haseley N, Hung DT. Eradication of bacterial persisters with antibiotic-generated hydroxyl radicals. *Proc Natl Acad Sci USA* 2012; **109**: 12147–12152.
94 Morones-Ramirez JR, Winkler JA, Spina CS, Collins JJ. Silver enhances antibiotic activity against gram-negative bacteria. *Sci Transl Med* 2013; **5**: 190ra81.
95 Kim JS, Heo P, Yang TJ, *et al*. Selective killing of bacterial persisters by a single chemical compound without affecting normal antibiotic-sensitive cells. *Antimicrob Agents Chemother* 2011; **55**: 5380–5383.
96 Pan J, Bahar AA, Syed H, Ren D. Reverting antibiotic tolerance of Pseudomonas aeruginosa PAO1 persister cells by (Z)-4-bromo-5-(bromomethylene)-3-methylfuran-2(5H)-one. *PLoS ONE* 2012; **7**: e45778.
97 Aagaard C, Hoang T, Dietrich J, *et al*. A multistage tuberculosis vaccine that confers efficient protection before and after exposure. *Nat Med* 2011; **17**: 189–194.
98 Wang CC, Zhu B, Fan X, Gicquel B, Zhang Y. Systems approach to tuberculosis vaccine development. *Respirology* 2013; **18**: 412–420.
99 Lowrie DB, Tascon RE, Bonato VL *et al*. Therapy of tuberculosis in mice by DNA vaccination. *Nature* 1999; **400**: 269–271.
100 van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 2009; **6**: 767–772.
101 LaFleur MD, Kumamoto CA, Lewis K. Candida albicans biofilms produce antifungal-tolerant persister cells. *Antimicrob Agents Chemother* 2006; **50**: 3839–3846.
102 Cheng Q, Kyle DE, Gatton ML. Artemisinin resistance in *Plasmodium falciparum*: A process linked to dormancy? *Int J Parasitol Drugs Drug Resist* 2012; **2**: 249–255.
103 Gupta PB, Chaffer CL, Weinberg RA. Cancer stem cells: mirage or reality? *Nat Med* 2009; **15**: 1010–1012.
104 Gupta PB, Onder TT, Jiang G *et al*. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 2009; **138**: 645–659.
105 Zhou JB, Zhang Y. Preclinical development of cancer stem cell drugs. *Expert Opin Drug Discov* 2009; **4**: 741–752.
106 McCune RM Jr, McDermott W, Tompsett R. The fate of *Mycobacterium tuberculosis* in mouse tissues as determined by the microbial enumeration technique. II. The conversion of tuberculous infection to the latent state by the administration of pyrazinamide and a companion drug. *J Exp Med* 1956; **104**: 763–802.
107 Andries K, Verhasselt P, Guillemont J *et al*. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* 2005; **307**: 223–227.
108 Fox W, Ellard GA, Mitchison DA. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946–1986, with relevant subsequent publications. *Int J Tuberc Lung Dis* 1999; **3**: S231–S279.
109 The World Health Organization. *Treatment of tuberculosis: guidelines*. 4th ed. Geneva: WHO, 2010.
110 Ryan FJ. Bacterial mutation in a stationary phase and the question of cell turnover. *J Gen Microbiol* 1959; **21**: 530–549.
111 Zhou J, Zhang H, Gu P, Bai J, Margolick JB, Zhang Y. NF-kappaB pathway inhibitors preferentially inhibit breast cancer stem-like cells. *Breast Cancer Res Treat* 2008; **111**: 419–427.
112 Zhou J, Zhang H, Gu P, Margolick JB, Yin D, Zhang Y. Cancer stem/progenitor cell active compound 8-quinolinol in combination with paclitaxel achieves an improved cure of breast cancer in the mouse model. *Breast Cancer Res Treat* 2009; **115**: 269–277.
113 Yuan S, Wang F, Chen G *et al*. Effective elimination of cancer stem cells by a novel drug combination strategy. *Stem Cells* 2013; **31**: 23–34.
114 Tashiro Y, Kawata K, Taniuchi A, Kakinuma K, May T, Okabe S. RelE-mediated dormancy is enhanced at high cell density in *Escherichia coli*. *J Bacteriol* 2012; **194**: 1169–1176.
115 Kim Y, Wood TK. Toxins Hha and CspD and small RNA regulator Hfq are involved in persister cell formation through MqsR in *Escherichia coli*. *Biochem Biophys Res Commun* 2009; **391**: 209–213.
116 Deb C, Lee CM, Dubey VS *et al*. A novel *in vitro* multiple-stress dormancy model for *Mycobacterium tuberculosis* generates a lipid-loaded, drug-tolerant, dormant pathogen. *PLoS ONE* 2009; **4**: e6077.
117 Sureka K, Ghosh B, Dasgupta A, Basu J, Kundu M, Bose I. Positive feedback and noise activate the stringent response regulator rel in mycobacteria. *PLoS ONE* 2008; **3**: e1771.
118 Debbia EA, Roveta S, Schito AM, Gualco L, Marchese A. Antibiotic persistence: the role of spontaneous DNA repair response. *Microb Drug Resist* 2001; **7**: 335–342.
119 Dorr T, Lewis K, Vulic M. SOS response induces persistence to fluoroquinolones in *Escherichia coli*. *PLoS Genet* 2009; **5**: e1000760.
120 Almiron M, Link AJ, Furlong D, Kolter R. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* 1992; **6**: 2646–2654.
121 Pandey AK, Sassetti CM. Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci USA* 2008; **105**: 4376–4380.

122 Adams KN, Takaki K, Connolly LE *et al*. Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell* 2011; **145**: 39–53.
123 Wang C, Mao Y, Yu J *et al*. PhoY2 of mycobacteria is required for metabolic homeostasis and stress response. *J Bacteriol* 2013; **195**: 243–252.
124 Moker N, Dean CR, Tao J. Pseudomonas aeruginosa increases formation of multidrug-tolerant persister cells in response to quorum-sensing signaling molecules. *J Bacteriol* 2010; **192**: 1946–1955.
125 Vega NM, Allison KR, Khalil AS, Collins JJ. Signaling-mediated bacterial persister formation. *Nat Chem Biol* 2012; **8**: 431–433.